Biochemical and Structural Characterization

of

Cation Chloride Cotransporters

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Stefan Warmuth
aus
Deutschland

Promotionskomitee

Prof. Raimund Dutzler (Leitung und Vorsitz)
Prof. Ben Schuler
Prof. Wolfram Welte

Zürich 2009

Acknowledgements

I would like to thank all the people who helped me realizing this thesis.

First of all I would like to thank Prof. Raimund Dutzler for giving me the opportunity to work on this challenging project with an excellent scientific environment. Furthermore I would like to thank Prof. Ben Schuler and Prof. Wolfram Welte for ideas, interesting discussions and completing my thesis committee.

I am grateful to the members of the laboratory for their support and discussions. Especially I would like to thank Sebastian for a great time in and outside the lab. I also would like to express my gratitude to Eric who opened a new window of hope in the project of CCCs and Eric and Sandra for being a super lab crew. Additionally I thank Laura & Iwan for performing great progress for the project during their diploma works.

Special thanks to Beat Blattmann and his coworkers at the NCCR crystallization platform for perfect support and to the whole institute.

Work is only productive with phases of relaxation. The coffee crowd and lunch bunch performed a great job in discussing everything inside and outside university.

My parents have always supported, understood and motivated me during all my educational time. Thank you very much!

Marlen, I would like to thank you for your encouragement through all the years and cheering me up every day!

Contents

Acknowledgements ..... **I**

List of Figures ..... **V**

List of Tables ..... **VII**

List of Abbreviations ..... **IX**

Zusammenfassung ..... **XI**

Summary ..... **XIII**

1 Introduction **1**

1.1 Membrane Transport Proteins 1

1.1.1 Membranes 1
1.1.2 Transporters and Channels 4

1.2 Cation Chloride Cotransporters 6

1.2.1 Overview of the Family 6
1.2.2 Physiology and Pharmacology 7
1.2.3 Structure and Function 10
1.2.4 Crystallization and Structure Solution of Membrane Proteins 15
1.2.5 Motivation 17

2 Results **18**

2.1 Overview 18

2.2 Homology Search 20

2.3 Prokaryotic Full-Length CCC 23

2.3.1 Cloning 23
2.3.2 Expression Tests 25
2.3.3 Purification of Prokaryotic CCCs 38
2.3.4 Crystallization 54
2.3.5 Materials and Methods 55

2.4 C-terminal Domain of CCCs 58

2.4.1 *Methanosarcina acetivorans* C-terminal Domain 59
2.4.2 Materials and Methods 84

3 General Discussion and Outlook **87**

4 Appendix 91
4.1 Membrane Domain of Prokaryotic CCCs 91
4.1.1 MaCCC Membrane Domain 91
4.2 Antibody Production 94
4.2.1 Binding Molecules that Facilitate Crystallization 94
4.2.2 Production and Identification of Positive Hybridoma Cell Lines 97
4.2.3 Fab Production and Purification 99
4.2.4 Fab Binding Analysis 100
4.2.5 Materials and Methods 102
5 Bibliography 105
Curriculum Vitae 111

List of Figures

Fig 1-1 The Membrane Potential 2
Fig 1-2 Goldman-Hodgkin-Katz Constant Field Equation 3
Fig 1-3 Transporters and Channels 4
Fig 1-4 Secondary Active Transporters 5
Fig 1-5 The Three Groups of Cation Chloride Cotransporters (CCCs) 6
Fig 1-6 Phylogenetic Tree and Sequence Alignment of Human CCC 7
Fig 1-7 Role of NCCs in the Transepithelial Movement of Ions 8
Fig 1-8 Role of NKCC2 in Transepithelial Ion Movement 9
Fig 1-9 Role of NKCC1 in Transepithelial Ion Transport 9
Fig 1-10 Simplified Sheme of the Topology of NKCCs 10
Fig 1-11 Splice Variants of NKCC2 12
Fig 1-12 Distribution of Ion Specificity on Transmembrane Segments of NKCC1 13
Fig 1-13 Sequence Alignment Between Human and Prokaryotic CCCs 15
Fig 1-14 Characteristics of Amphiphiles 16
Fig 2-1 Pipeline Towards the Structure Solution of a Membrane Protein 18
Fig 2-2 Detergents 19
Fig 2-3 Sequence Alignment of Selected Prokaryotic Homologues 22
Fig 2-4 FX Cloning 24
Fig 2-5 Expression Constructs for Prokaryotic Homologues 25
Fig 2-6 GFP as Expression and Folding Indicator 27
Fig 2-7 Analysis of Expression in *E. coli* Using GFP Folding Assay 28
Fig 2-8 Fine Tuning of Expression in *E. coli* 29
Fig 2-9 Expression Test in *E. coli* with N-terminal His-tag 30
Fig 2-10 Expression of CCCs in *E. coli* with C-terminal His-tag Followed by Detergent Sedimentation Assay 30
Fig 2-11 Expression Test in *E. coli* Using Different Media 33
Fig 2-12 Analysis of Expression in *L. lactis* with In-Gel GFP Fluorescence 35
Fig 2-13 Analysis of Expression in *L. lactis* with Immunostaining 36
Fig 2-14 3C Protease Site 38
Fig 2-15 Calibration Curve of S200 GE Healthcare Column 39
Fig 2-16 SfCCC Extraction Tests in Dependence of Expression Conditions 40
Fig 2-17 SfCCC SEC in Fos-cholin 12 41
Fig 2-18 SfCCC Detergent Stability Screen 42
Fig 2-19 SynCCC SEC Behavior and Stability Test in Fos-cholin 12 43
Fig 2-20 SynCCC SEC Behavior and Stability 44

Fig 2-21 MaCCC-His_6 IMAC Purification Test 45
Fig 2-22 MaCCC-His_{10} IMAC Purification Test 46
Fig 2-23 MaCCC SEC Profile....47
Fig 2-24 MaCCC Purification in Fos-cholin 12....48
Fig 2-25 MaCCC Detergent Stability Test....49
Fig 2-26 MaCCC-10His SEC in DM+Lipids 50
Fig 2-27 MaCCC Cysteine Crosslink 51
Fig 2-28 SEC-MALS of Oxidized MaCCC E505Q622C....52
Fig 2-29 MALS of Reduced MaCCC E505Q622C 53
Fig 2-30 Expression Test of Prokaryotic C-Terminal Domains 59
Fig 2-31 MaCCC C-Terminal Domain Construct Design 60
Fig 2-32 MaCCC Cdom Purification 61
Fig 2-33 Analytical Ultracentrifugation of MaCCC Cdom 62
Fig 2-34 MaCCC Cdom Se-Methionine Derivatization 63
Fig 2-35 MaCCC Cdom SAD Data Collection Statistics 63
Fig 2-36 Heavy Atom site Occupancy After Heavy Atom Search in SHELXD 64
Fig 2-37 Electron Density of MaCCC C-Terminal Domain....65
Fig 2-38 Density Modification in SHELXE 65
Fig 2-39 Crystal Contact in Crystal Form $P3_221$ 66
Fig 3-1 Model for CCC Organization 88
Fig 4-1 MaCCC Amino Acid Sequence of the Linker Between the Membrane Domain and the C-Terminal Domain 91
Fig 4-2 MaCCC Membrane Domain Expression Test....92
Fig 4-3 Expression Test of MaCCC Membrane Domain....93
Fig 4-4 MaCCC Membrane Domain F-HPLC....94
Fig 4-5 Antibody Molecules and Fragments....95
Fig 4-6 Shematic Representation of DARPins....96
Fig 4-7 Co-Crystallization with Proteinaceous Binders....96
Fig 4-8 Antibody Binding Against MaCCC Full-Length and C-Terminal Domain....98
Fig 4-9 Epitope Characterization of Selected Monoclonal Antibodies 99
Fig 4-10 Purification and Binding of SW27 Fab 101
Fig 4-11 Anti-MaCC Fab Binding Test 101

Fig S-1 Unstructured regions in human NKCC2. 78
Fig S-2 Sequence conservation in the C-terminal domain....79
Fig S-3 Experimental electron density....79
Fig S-4 Structural homology to USP domain proteins....80
Fig S-5 Analytical ultracentrifugation of the MaCCC domain 81
Fig S-6 MaCCC domain 'head to tail' dimer 81
Fig S-7 Crystal packing....82
Fig S-8 Structure of the 'head to head' dimer 83

List of Tables

Tab 2-1 Prokaryotic Homologues 21
Tab 2-2 Summary of Expression Tests in *E. coli* Using the Arabinose Induction System 31
Tab 2-3 Overview of Expression Using T7 Promoter in *E. coli* 34
Tab 2-4 Overview of the Expression Tests in *L. Lactis* 36
Tab 2-5 Summary of the Expression Test in *E. coli* and *L. Lactis* 37
Tab 2-6 Crystal Forms of MaCCC C-Terminal Domain 66
Tab 2-7 Protease Sites and Their Overhangs 67
Tab 4-1 Fab Binding to MaCCC 102

List of Abbreviations

ATCC	American Type Culture Collection
CAP	Catabolic Activator Protein
CCC	Cation Chloride Cotransporters
DCT	Distal Convoluted Tubule
DDM	Dodecyl Maltoside
DIAO	Dihydroindenoyloxy-Alkanoic acid
DM	Decyl Maltoside
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
ESI	Electron Spray Ionization
FX	Fragment Exchange
GFP	Green Fluorescent Protein
HAT	Hypoxanthine Aminopterin Thymidine
HGPRT	Hypoxanthine-Guanine Phosphoribosyltransferase
HRP	Horseradish Peroxidase
IDA	Iminodiacetic Acid
IEC	Ion Exchange Chromatography
IMAC	Immobilized Metal Affinity Chromatography
KCC	Potassium Chloride Cotransporter
LB	Lysogeny Broth
LDAO	Lauryl Dimethylamine-N-Oxide
MALDI	Matrix-Assisted Laser Desorption/Ionization
MALS	Multi Angle Light Scattering
MS	Mass Spectroscopy
M_w	Molecular Weight
MWCO	Molecular Weight Cut Off
NCC	Sodium Chloride Cotransporters
NKCC	Sodium Potassium Chloride Cotransporter
NTA	Nitrilotriacetic Acid
OG	Octyl Glucoside
PAGE	Polyacrylamid Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCC	Pasteur Culture Collection of Cyanoacteria
PCR	Polymerase Chain Reaction
PDC	Protein Detergent Complex
PVDF	Poly Vinyl Diflourid
RI	Refractive Index
SDS	Sodium Dodecyl Sulfate
SEC	Size Exclusion Chromatography
TALH	Thick Ascending Loop of Henle
TCEP	Tris(2-carboxyethyl)phosphine
TB	Terrific Broth

TOF	Time of Flight

Zusammenfassung

Die Kation Chlorid Cotransporter (CCCs) stellen eine wichtige Klasse sekundär aktiver Membrantransportproteine dar, die in allen Organen des menschlichen Körpers vorkommen. Sie katalysieren den elektroneutralen Kation Cl^- Symport durch die Plasmamembran. CCCs spielen eine wichtige Rolle in der Regulation des Zellvolumens, beim Transport von Ionen durch Epithelschichten und in der Regulation intrazellulärer Cl^- -Konzentrationen. Im Laufe der letzten zwei Jahrzehnte wurden sieben humane Isoformen identifiziert und gemäss dem Kation, welches an den Cl^- Transport gekoppelt ist, klassifiziert. Die Erforschung von Transporteigenschaften der humanen Transporter in HEK-293 Zellen ermöglichte die Messung von Ionenaffinitäten und die Identifizierung von Inhibitoren. Zusätzlich zeigten biochemische Experimente die dimere Organisation der NKCCs und die Bedeutung der C-terminalen Domäne für die Homooligomerisation. Obwohl die physiologische Rolle der verschiedenen Transporter bereits im Detail erforscht worden ist, gibt es bis heute keine strukturellen Informationen zu dieser Transporterfamilie.

Im Rahmen dieser Dissertation wurden mehrere prokaryotische Homologe der Kation-Chlorid-Cotransporterfamilie identifiziert und deren Expression, Aufreinigung und Kristallisationsverhalten untersucht, um die erste CCC Struktur mit Hilfe der Röntgenstrukturkristallographie zu bestimmen. Das gesamte Protein und die isolierten C-terminalen Domänen dieser Homologe wurden kloniert und deren Expression unter verschiedenen Bedingungen getestet. Die Überexpression des Proteins wurde durch den T7- und P_{BAD} Promoter in *E.coli* reguliert und für die meisten Proteine konnte dadurch eine Expression herbeigeführt werden. Eine GFP Fusion am C-Terminus wurde als Faltungsindikator verwendet, um frühzeitig Auskunft über die Qualität des Proteins zu erhalten. Nach der Extraktion des Transporters in Detergenzien konnten nur fünf Homologe mit ausreichender Proteinmenge aufgereinigt werden, die weitere biochemische Untersuchungen erlaubte. Obwohl vier der Transporter eine Tendenz zur Aggregation zeigten, nachdem sie aus der Membran gelöst wurden, konnte ein prokaryotisches Homolog aus der Familie der Archäen, *Methanosarcina acetivorans C2A*, MaCCC, vollständig in verschiedenen Detergenzien aufgereinigt werden. Dieses Konstrukt bildet ein wichtiges Modelsystem für zukünftige strukturelle und funktionelle Untersuchungen dieser Proteinfamilie.

Um Einblicke in die strukturelle Organisation der CCC-Transporter zu erhalten, konnte während dieser Dissertation die Struktur der cytosolischen Domäne von MaCCC mit einer Auflösung von 1.9 Å bestimmt werden. Diese Struktur zeigt einen neuen Aufbau einer regulatorischen Einheit von Transportproteinen, und hat Ähnlichkeit mit

dem Aufbau von universellen Stressproteinen. Da die Quartärstruktur des Proteins in Kristallen nicht eindeutig war, wurden biochemische Methoden verwendet, um die oligomere Organisation zu charakterisieren. Die analytische Ultrazentrifugation und die Grössenausschlusschromatographie haben gezeigt, dass die Domäne in Lösung Dimere formt. Ein mögliches Dimer Interface, welches in den Kristallen beobachtet wurde, wurde durch zielgerichtete Cystein-Mutanten bestätigt, die in der Lage waren, Disulfidbrücken zu bilden. Da der gesamte Transporter ebenso Dimere bildet, konnte mit denselben Crosslinks, die Relevanz der beobachteten Interaktionen für das gesamte Protein gezeigt werden. Die Experimente liefern erste strukturelle Einblicke in die Familie der CCCs und etablieren die dimere Organisation der CCC Transporter.

Summary

The Cation chloride cotransporters (CCCs) constitute an important class of secondary active membrane transport proteins that are ubiquitously expressed in human. They catalyze the electro-neutral cation-Cl^- symport across the plasma membrane. CCCs play an important role in cell volume regulation, the transepithelial movement of ions and the regulation of the intracellular Cl^- concentration (Gamba 2005). Seven human isoforms were identified over the last decade and classified according to the cation coupled to Cl^- transport. The investigation of transport characteristics of certain human transporters expressed in HEK-293 cells allowed the measurement of ion affinities and the identification of inhibitors. Additional biochemical experiments proposed the dimeric organization of NKCCs and the important role of the C-terminal domain for homo-oligomerization . While the physiological roles of different transporters were investigated in detail, the family still lacks structural representation.

In the course of this thesis several prokaryotic homologues of the cation chloride cotransporter family have been identified and were investigated with respect to their expression, purification and crystallization behavior with the goal to determine the first CCC structure by X-ray crystallography. The full-length protein and the isolated C-terminal domain of these homologues were cloned and tested for expression under different conditions. Protein overexpression was investigated in *E. coli* under the control of the T7- and the P_{BAD} promoters. Under these conditions most proteins showed detectable expression. The addition of a C-terminal GFP fusion was used as folding indicator to determine the quality of the protein at an early stage. After extraction of the full-length transporter in detergents and purification, only five homologues provided sufficient protein for a biochemical characterization. While four of the transporters showed the tendency to aggregate once they were solubilized from the membrane, one prokaryotic homologue from the archaea *Methanosarcina acetivorans C2A* , MaCCC, allowed the purification of the full-length protein in different detergents. This construct provides an important model system for future structural and functional investigations.

To gain insight into the structural organization of CCC transporters the structure of the cytosolic domain of MaCCC was determined in parallel at 1.9 Å resolution. This structure revealed the architecture of a novel fold for a regulatory unit of a transport protein that is distantly related to universal stress proteins. Since the quaternary structure of the protein in the crystals was ambiguous, biochemical methods were used to characterize the oligomeric organization. Analytical ultracentrifugation and size exclusion chromatography revealed that the domain forms dimers in solution. A possible dimer interface observed in the crystals was confirmed by the site-directed

introduction of cysteine mutants that were able to form disulfide bonds. Since the full-length transporter also forms dimers, the relevance of the observed interactions was confirmed by forming the same cysteine crosslink in the context of the full-length transporter. The experiments have thus provided a first structural insight into the CCC family and have established the dimeric organization of CCC transporters.

1 Introduction

1.1 Membrane Transport Proteins

1.1.1 Membranes

Membranes surround all living cells and their intracellular organelles. They are the direct interface between the inside and outside of the respective compartments. Membranes mainly consist of amphiphilic phospholipids. These molecules contain a hydrophilic head group of glycerol esters, a phosphate group, and hydrophobic tails of two fatty acid side chains. In water phospholipids spontaneously form lipid bilayers, where the head groups are pointed towards the water and the tails form a hydrophobic core. This core has a dielectric constant of about $\varepsilon=2$ which is drastically lower than the dielectric constant in water ($\varepsilon=80$). This property determines the permeability properties of the membranes. Non-polar molecules and small-polar molecules like water can pass the membrane to different extend, whereas charged ions or large-polar macromolecules such as proteins and nucleic acids are excluded from passage. Due to these dielectric features membranes behave like electric capacitors, a property that is used for the creation and conservation of the membrane potential.

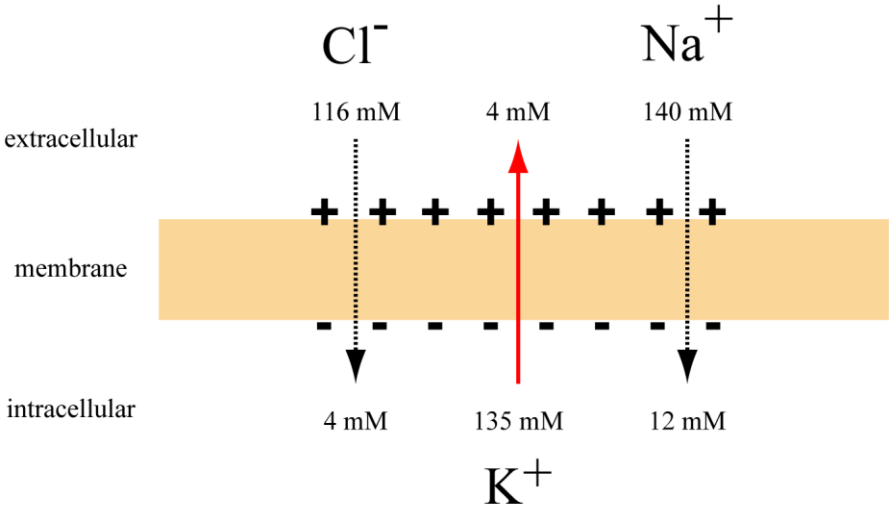
Membrane Proteins are associated with the lipid bilayer and functionalize the membranes. They have important functions in cell adhesion as well as for energy storage and they catalyze the specific transport of signals, metabolites, ions, proteins and other solutes across the membrane. Membrane proteins may be attached to the inner or outer surface or penetrate both leaflets of the membrane. The amphiphilic character of membrane proteins is characterized by a hydrophilic surface that is in contact with the aqueous solution or the polar headgroups of the lipids, and a hydrophobic belt that interacts with alkyl side chains of the lipids in the core of the membrane.

Membrane Transport is dependent on a driving force that fuels the transport process. Passive transport or facilitated diffusion uses an existing concentration gradient. Solutes are passing the membrane to balance the concentration on both sides of the membrane. Active transport is an energy-dependent process that moves solutes across the membranes against their concentration gradient. Primary active transport directly uses chemical energy like hydrolysis of ATP for the transport of solutes and ions. Transported ions energize the membrane through an asymmetric distribution of charge and concentration on both sides of the membrane resulting in an electrochemical gradient. In cells this gradient is usually built up with protons (H^+) resulting in a proton motive force or with sodium ions (Na^+) creating the sodium motive force. Secondary active transporters use the energy stored in the ion gradient across the membrane to

transport solutes and ions against their concentration gradient. In animals, electrochemical gradients across the cytoplasmic membrane are due to the asymmetric distribution of Na^+ and potassium ions (K^+).

The **Membrane Potential** of an animal cell is dominated by the three most abundant ions: K^+, Na^+ and chloride (Cl^-). The controlled movement of these ions across the membrane underlies such important processes as electric signaling as well as the maintenance of the osmotic balance, and it provides energy for secondary active transporters. The membrane potential originates from two features: large concentration gradients for K^+ and Na^+ across the membrane and the selective permeability of the

A



B

$$\mu_x = -\frac{RT}{zF} \ln \frac{[X]_{in}}{[X]_{out}}$$

C

Ion	[intracellular], mM	[extracellular], mM	μ_x, mV
K^+	135	4	-92
Na^+	12	140	+64
Cl^-	4	116	-88

Fig 1-1 The Membrane Potential
A Na^+, K^+ and Cl^- are asymmetrically distributed on both sides of the membrane. Na^+ and Cl^- are enriched in the extracellular space and K^+ in the intracellular space. While Cl^- and Na^+ have limited permeability across the membrane (dashed arrows) K^+ can move across the membrane through open K^+ channels (red arrow) creating a positive charge outside and a negative charge inside. **B** The resulting chemical potential of the ions can be calculated using the Nernst equation. μ_X, electrochemical potential for ion X; R, gas constant; T, temperature; z, charge; F, Faraday constant; $[X]_{in}$, concentration for ion X inside; $[X]_{out}$, concentration for ion X outside. **C** For every ion X (K^+, Na^+, Cl^-) a specific electrochemical potential can be calculated.

membrane for those ions mediated by selective-gated ion channels. The electrochemical potential under equilibrium conditions is a balance between the concentration-dependent chemical potential and the permeability-dependent electrical potential.

The chemical potential over the membrane derives from the concentration differences of K^+ (135 mM inside and 4 mM outside the cell) and Na^+ (12 mM inside and 140 mM outside the cell) (Fig 1-1A). The electrical potential results from the selective permeability of the membrane for certain ions. When K^+ crosses the membrane through constantly open potassium channels following its concentration gradient, the negative counter ion remains intracellular. The resulting negatively charged inside and the positively charged outside of the cell creates an electrical force opposing the chemical force (Fig 1-1A). A net flux of K^+ ions out of the cell will continue as long as the chemical potential is higher than the electrical potential. Under equilibrium conditions the chemical potential balances the electrical potential and can be described by the Nernst equation (Fig 1-1B) resulting in a defined electrochemical potential for each ion (Fig 1-1C).

The transmembrane potential is depended on all ions and can be estimated considering the selective permeability of the membrane for each of the three major ions and their concentration distribution. The Goldman-Hodgkin-Katz Constant Field equation (Fig 1-2) describes the steady state potential taking all ion fluxes into account.

$$V_m \cong -\frac{RT}{zF} \ln \frac{P_{K^+}[K^+]_{in} + P_{Na^+}[Na^+]_{in} + P_{Cl^-}[Cl^-]_{out}}{P_{K^+}[K^+]_{out} + P_{Na^+}[Na^+]_{out} + P_{Cl^-}[Cl^-]_{in}}$$

with

$$P_{K^+} = 10^{-7}, P_{Na^+} = 10^{-9}, P_{Cl^-} = 10^{-8}$$

$$V_m = -84mV$$

Fig 1-2 Goldman-Hodgkin-Katz Constant Field Equation
The Goldman equation describes the membrane potential (V_m) using the concentration differences of the ions inside and outside of the cell ($[X]_{in}$ and $[X]_{out}$, respectively (with $X=K^+,Na^+$ andCl^-) and their selective permeability through the membrane (P_X). R, gas constant; T, temperature; z, charge; F, faraday constant.

Since K^+ ions are highly permeable (10^{-7}) for resting cells they dominate the electrochemical membrane potential, while Na^+ ions have a two order of magnitude lower permeability (10^{-9}).

The resulting negatively charged interior of animal cells with respect to the external solution has an influence on the ion motive forces of Na^+ and K^+. Na^+ following its chemical and electrical gradient exerts a higher force than K^+ following its higher chemical gradient but moving against the electrical gradient. Thus, the Na^+ motive force is more powerful and mostly used for energy-dependent transport processes.

1.1.2 Transporters and Channels

Membrane transport proteins provide solutes with hydrophilic interactions to mimic the dielectric environment of water during transport across the membrane. Transporters and channels are the main classes of membrane ion transport proteins that can be distinguished by their mode of action.

Transporters can use the energy from ATP hydrolysis or light (primary active transport) or the energy stores in an ion gradient (secondary active transport) to transport ions and solutes against their concentration gradient. During this catalytic process they work by an alternate access mechanism that requires at least two distinct protein conformations (Fig 1-3A). The two conformations expose substrate-binding sites to either site of the membrane and often have a high affinity for the cargo on one side of the membrane and a low affinity for the cargo at the other side of the membrane. In some cases, a defined occluded state is observed that represents the activated conformation during transport catalysis where the cargo is completely enclosed by the transporter.

Channels facilitate diffusion of ions across the membrane along their chemical or electrochemical gradient. They consist of a transmembrane domain that contains the ion selectivity filter of the channel and a gate that prevents ion conduction through the channel in the closed state (Fig 1-3B). Channels can be ligand-gated, mechanically-gated, voltage-gated or light-gated.

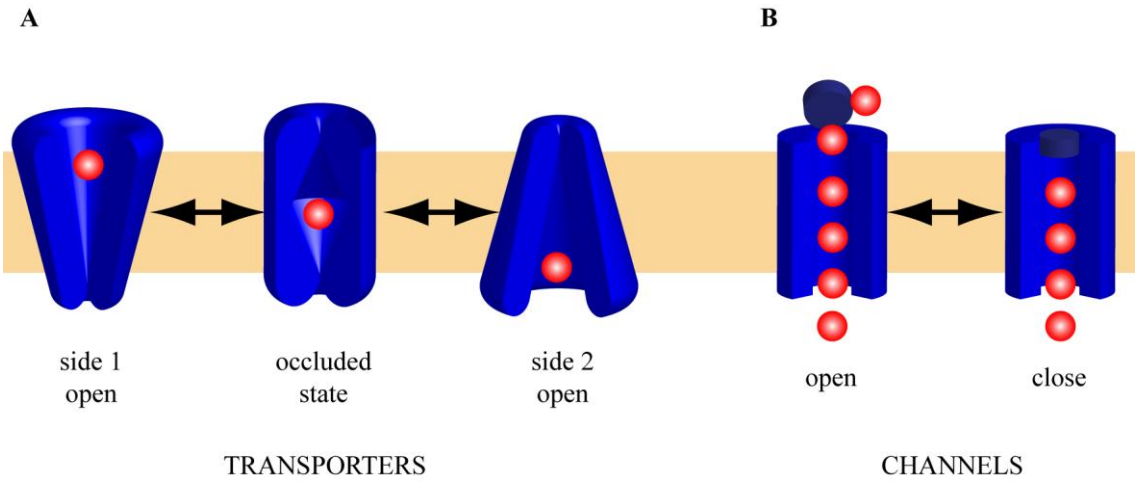


Fig 1-3 Transporters and Channels
Transporters and channels are the two most important classes of membrane transport proteins. They are classified based on their mode of action. **(A)** Transporters follow the alternate access mechanism with two distinct states exposing binding sites to either side of the membrane that have different affinities, high and low, for the cargo. A third occluded state encloses the cargo completely. **(B)** Channels are regulated by a gate to open and close the channel.

Primary Active Transport is accomplished by membrane bound ATPases and by ATP binding cassette (ABC) transporters. Membrane bound ATPases translocate ions across the membrane and in that way build up of ion gradients for secondary active transport processes. Important members are the Na^+/K^+ ATPase (in the plasma

membrane) (Morth, Pedersen et al. 2007), the H^+/K^+ ATPase (which is important in gastric acidification) and the Ca^{2+} ATPase (in the sarcoplasmatic reticulum of muscle cells) (Olesen, Picard et al. 2007).

The Na^+/K^+ ATPase catalyzes the transport of $3Na^+$ outside and $2K^+$ inside the cell for each cycle of ATP hydrolysis. This results in an asymmetric ion distribution of Na^+ (12 mM inside to 140 mM outside the cell) and K^+ (135 mM inside to 4 mM outside the cell) that is the basis for the membrane potential (see above).

ABC transporters are driven by ATP hydrolysis on the inside of the cell to import (prokaryotes) or export (eukaryotes) their cargo. Recently, structures of all states during the transport cycle were elucidated and theories were formulated how the ABC motor domains convert the chemical energy into defined transport motions (Oldham, Davidson et al. 2008).

Secondary Active Transporters couple the transport of one ion along its electrochemical potential with the translocation of the substrate against its concentration gradient across the membrane. They either work as symporters, where both molecules pass through the membrane in the same direction such as the Na^+/glucose transporters (Abramson and Wright 2009) (Fig 1-4A), or antiporters such as H^+/Cl^- transporters (Dutzler, Campbell et al. 2002) (Accardi and Miller 2004), where the downhill transport of ions in one direction provides energy for the active transport of the substrate in the opposite direction (Fig 1-4B).

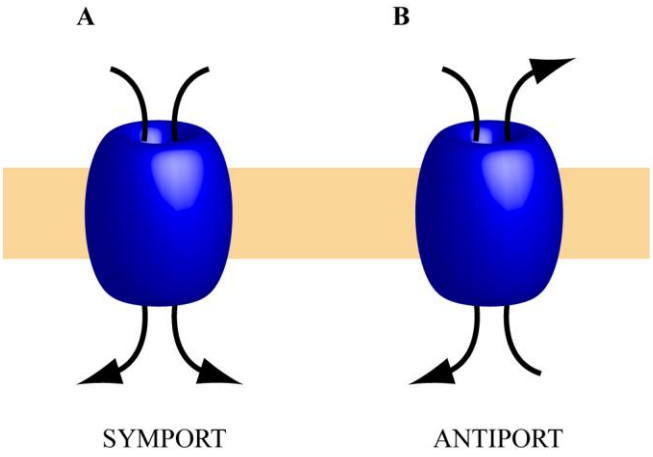


Fig 1-4 Secondary Active Transporters
Secondary active transporters couple the transport of energized ion in two ways. **(A)** Both move in the same direction through the membrane like in symporters or **(B)** in opposite directions as in antiporters.

1.2 Cation Chloride Cotransporters

1.2.1 Overview of the Family

Cation Chloride Cotransporters (CCCs) are found in all kingdoms of life where they play an important role in different physiological functions. They perform an electro neutral cation-Cl^- coupled symport that is secondary active (Gamba 2005). CCCs are divided into three big groups based on the cation coupled to the Cl^- transport (Fig 1-5). Sodium chloride cotransporters (NCCs, Fig 1-5B) transport one Cl^- together with one Na^+ ion while potassium chloride cotransporters (KCCs, Fig 1-5C) couple one K^+ per Cl^- ion. The third group, called sodium potassium cotransporters (NKCCs, Fig 1-5A), transport all three ions, Na^+, K^+ and Cl^- at a ratio of 1:1:2. Based on their pharmacological profile four groups were defined:

- benzothiadiazine sensitive NCCs,
- sulfamoylbenzoic (bumetanide) sensitive NCCs,
- sulfamoylbenzoic (bumetanide) sensitive NKCCs and
- dihydroindenoyloxy-alkanoic acid (DIAO) sensitive KCCs.

The driving force of NKCCS and NCCs to transport ions inside the cell is the Na^+ motive force while KCCs use the K^+ motive force for transporting ions outside of the cell.

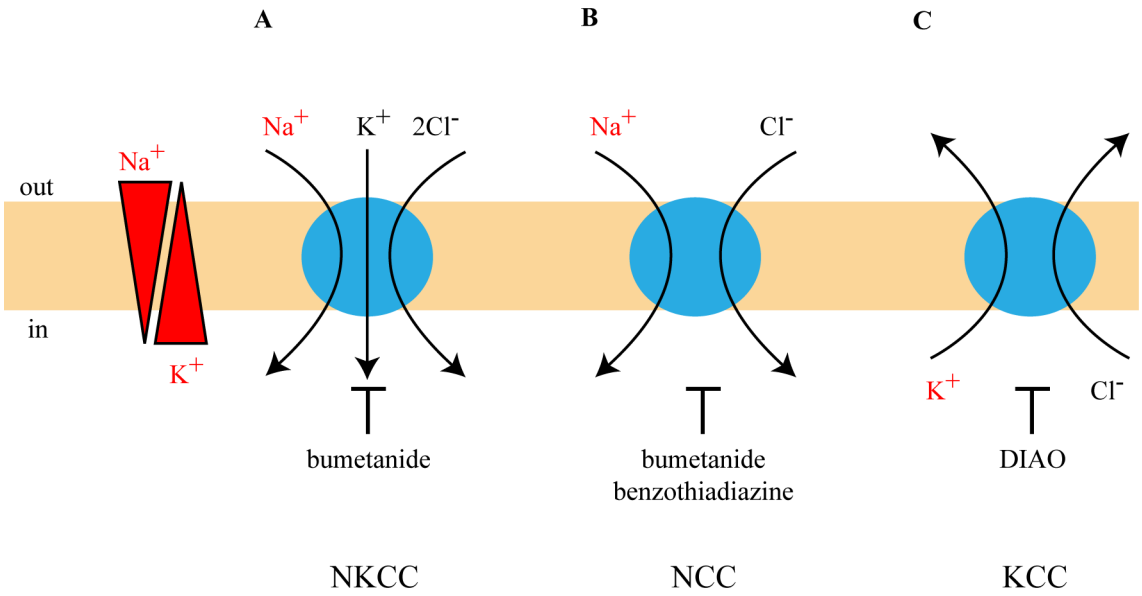


Fig 1-5 The Three Groups of Cation Chloride Cotransporters (CCCs)
CCCs are divided into three groups based on the coupled cation that is transported with the chloride ion: **A** Na^+-K^+-Cl^--cotransporter (NKCC), **B** Na^+-Cl^--cotransporter (NCC) and **C** K^+-Cl^--cotransporter (KCC). They use the Na^+ and K^+ electrochemical gradient for transport (red triangles) which also defines the direction of ion movement: Inside for NKCCs and NCCs and outside for KCCs. The net charge transported is zero. The transport can be inhibited by various compounds: bumetanide (NKCCs, NCCs), benzothiadiazine (NCCs) and DIAO (KCCs).

In human one NCC, four KCC and two NKCC isoforms were identified. Sequence comparisons of these isoforms reveal a high sequence identity between Na^+-driven cotransporters like NCC and NKCCs (around 50 %) with a lower sequence identity to K^+-driven cotransporters (around 20 % identity between NKCCs and KCCs). The KCCs

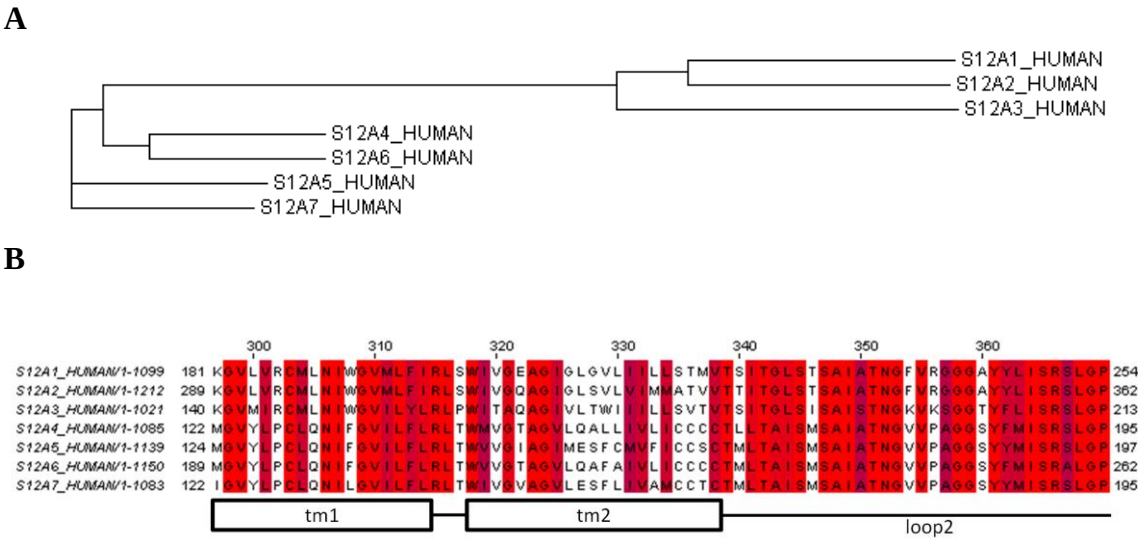


Fig 1-6 Phylogenetic Tree and Sequence Alignment of Human CCC
A Phylogenetic analysis reveals the clustering of Na^+-dependent CCCs (S12A1-NKCC2, S12A2-NKCC1, S12A3-NCC) and K^+-dependent CCCs. Within the Na^+ dependent CCCs, NKCCs (S12A1, S12A2) show a high sequence relationship. KCCs are subdivided into two branches consisting of KCC1 and KCC3 (S12A4, S12A6 respectively) and KCC2 and KCC4 (S12A5, S12A7 respectively). **B** The sequence alignment between the 7 human CCCs shows a high homology between the family members. Identical residues are colored in red, similar residues are colored in ruby. The topology prediction is indicated based on NKCC2 (S12A1) with putative transmembrane helices (tm) 1 and 2 and the intracellular connecting loop 2.

are clustered in pairs, KCC1 and KCC3 and KCC2 and KCC4 have highest mutual sequence conservation (Fig 1-6A and B). All isoforms show a distinct expression pattern. NKCC2 and NCC exist exclusively in the kidney and are apically expressed. NKCC1 is ubiquitously expressed on the basolateral side of the epithelial. KCC1 and 4 are also ubiquitously expressed, while KCC2 is exclusively expressed in the central nervous system (CNS). KCC3 plays an important role in the CNS but is also detected in the inner ear and in vascular tissues.

1.2.2 Physiology and Pharmacology

CCCs have a major impact on cell volume regulation, transepithelial movement of ions and the regulation of the intracellular Cl^- concentration.

Cell Volume Regulation is primarily an adaptation to the extracellular osmolarity by a variation of the internal osmolarity. The key players in this regulation are NKCCs and KCCs. If cells are exposed to a hypertonic environment they loose water and the cell shrinks. NKCCs transport Na^+, K^+ and Cl^- inside the cell and reduce the osmotic

differences to the extracellular space and thus recover the original cell volume. Since KCCs facilitate the transport of ions out of the cell they play an important role when the cell senses hypotonic conditions. The extrusion of KCl increases the water efflux and thus prevents cell swelling.

Transepithelial Movement of Ions is particularly important for the reabsorption of salt in the kidney. NCCs are expressed apically in the distal convoluted tubule (DCT) and NKCC2 apically in the thick ascending loop of Henle (TALH). Both proteins are major proteins involved in the translocation of ions from the lumen into the cells.

NCCs provide the major sodium reabsorption pathway along the DCT (Fig 1-7). They facilitate the transport of sodium chloride (NaCl) into the cell driven by the Na^+-concentration generated by the basolateral polarized Na^+/K^+ ATPase. The accumulated K^+ that enters the cell through the Na^+/K^+ ATPase is secreted to the lumen via apical KCCs and ROMK K^+ channels. This process also influences the Ca^{2+} reabsorption through Ca^{2+} channels in the apical membrane. If NCCs are blocked through thiazide-type diuretics an electrical gradient is created because the Na^+/K^+ ATPase continuously

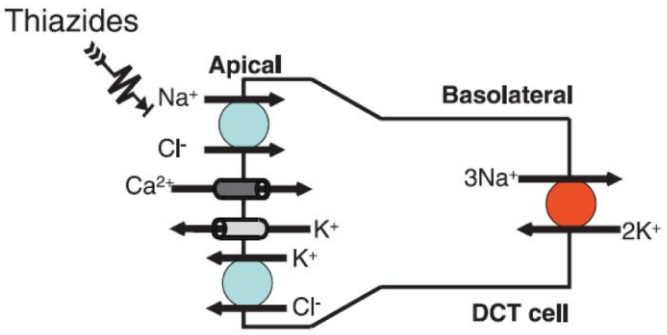


Fig 1-7 Role of NCCs in the Transepithelial Movement of Ions
The transepithelial salt movement in the distal convoluted tubule (DCT) is a team play between the Na^+/K^+ ATPase (red) that creates the Na^+ motive force and NCCs (blue, top), the major salt transport into the cell. KCCs (blue, bottom) facilitate the K^+ transport back into the lumen. The K^+ transport through K^+ ROMK channels (light grey) back into the lumen promotes the reabsorption of Ca^{2+} through Ca^{2+} channels (dark grey). Thiazides inhibit NCCs and stop reabsorption of ions from the lumen. (Gamba 2005)

reduces the Na^+ concentration in the cell. Thus, the entry of Ca^{2+} is promoted. The use of thiazides like chlorthalidone, hydrochlorothiazide, bendroflumethaizide and metolazone are used as drugs against hypertension and calcium stones and osteoporosis.

NKCC2 is the primary NaCl reabsorption pathway in the TALH that is responsible for the reabsorption of 15-20 % of the glomerular filtrate (Fig 1-8). The basolateral polarized Na^+/K^+ ATPase creates a membrane potential that drives the inward transport of Na^+, K^+ and Cl^- by NKCC2 from the lumen into the cell. Na^+ and Cl^- leave the cell basolaterally through the Na^+/K^+ ATPase and CLC-K, respectively. K^+ is transported back to the potassium-ion depleted tubular lumen via the ROMK K^+ channel to recycle K^+ for the NKCC2-dependent transport. During this transport cycle a positive charge

remains in the tubule originated from the extruded K^+ that drives the diffusion of one Na^+, Ca^{2+}or Mg^{2+} through a paracellular pathway across the tight junctions into the interstitial space. The major ammonia reabsorption in the TALH is facilitated by the replacement of K^+ in NKCC2 by ammonium resulting in a $Na^+/NH_4^+/Cl^-$ cotransport. The most potent loop diuretics like furosemide or bumetanide that target NKCC2 produce a significant natriuresis and diuresis, and are used against hypertensions and hypercalcemia.

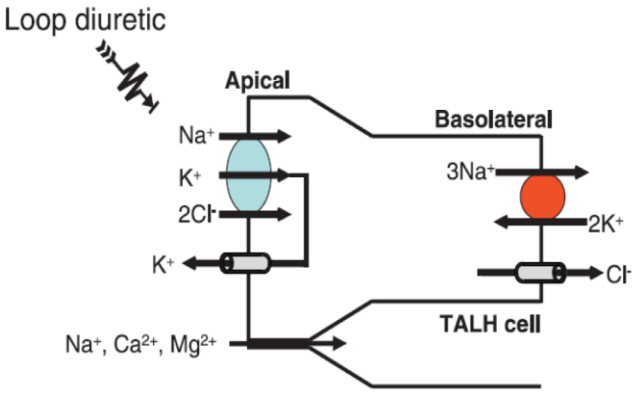


Fig 1-8 Role of NKCC2 in Transepithelial Ion Movement
Salt reabsorption in the thick ascending loop of henle (TALH) is driven by the concentration gradients created by the basolateral polarized Na^+/K^+ ATPase. NKCC2 (blue) facilitates the movement of Na^+, K^+ and Cl^- from the lumen into the cell. K^+ is recycled to the lumen via ROM-K K^+ channels (gray, apical). Cl^- is transported out of the cell via basolateral CLC-K channels (grey, basolateral). Other cations (Na^+, Ca^{2+}, Mg^{2+}) are reabsorbed into the interstitial space via tight junctions. Loop diuretics inhibit NKCC2 and deactivate reabsorption of ions from the lumen. (Gamba 2005)

NKCC1 is involved in transepithelial ion movement in epithelial secretory cells (Fig 1-9). Basolateral polarized NKCC1 provides the cells with Cl^- that is secreted through the apical membrane. Na^+ is recycled via the Na^+/K^+ ATPase and K^+ through a conductive pathway in the basolateral membrane. Cl^- in contrast is secreted into the

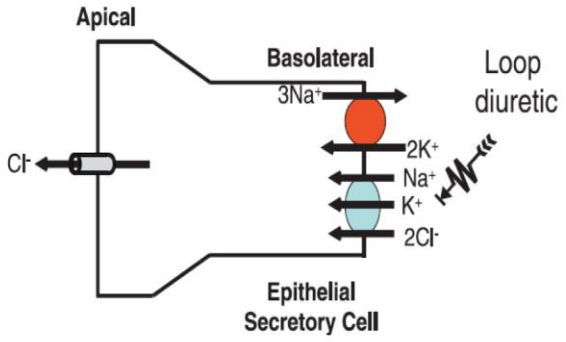


Fig 1-9 Role of NKCC1 in Transepithelial Ion Transport
In epithelial secretory cells the basolateral Na+/K+-ATPase (red) is creating a concentration gradient that fuels the uptake of salt via NKCC1 (blue) into the cell. Cl^- is secreted to the lumen via chloride channels (grey) *e.g.* the cystic fibrosis transmembrane conductance regulator (CFTR). (Gamba 2005)

lumen via several pathways. Cystic fibrosis transmembrane conductance regulator (CFTR) is the most important Cl^- dislocator in the apical membrane. This pathway is essential for the acidification of the stomach and in the maintenance of the mucosal fluid viscosity in the lung which is impaired in cystic fibrosis.

Regulation of the Intracellular Cl^- Concentration plays an important role in the regulation of γ-aminobutyric acid (GABA) and glycine mediated responses in neurons (Schomberg, Bauer et al. 2003). KCC2 that is exclusively expressed in the CNS defines together with NKCC1 the type and magnitude of responses to neurotransmitters in the development and maturation of neurons. GABA binding to the GABA receptor mediates the opening of a ligand-gated Cl^- channel. Depending on the Cl^- concentration in the neuron, the Cl^- channel opening results in a Cl^- efflux and depolarization or Cl^- influx and hyperpolarization. In immature neurons, the Cl^- concentration is above the electrochemical equilibrium due to high expression rates of NKCC1 and low expression rates of KCC2. Responses to GABA are thus excitory due to a depolarization of the membrane potential. Mature neurons, in contrast, contain a Cl^- concentration that is below its electrochemical equilibrium due to a downregulation of NKCC1 and a high expression of KCC2. This results in the influx of Cl^- upon to GABA receptor activation because of the hyperpolarization of the membrane.

1.2.3 Structure and Function

CCCs are 120-200 kDa big, homodimeric glycoproteins. They share a common architecture that is composed of a relatively small intracellular N-terminus (41-285 residues) followed by a transmembrane domain (~470 residues for Na^+-dependent CCCs to ~540 residues for K^+-dependent CCCs) and a large intracellular C-terminal

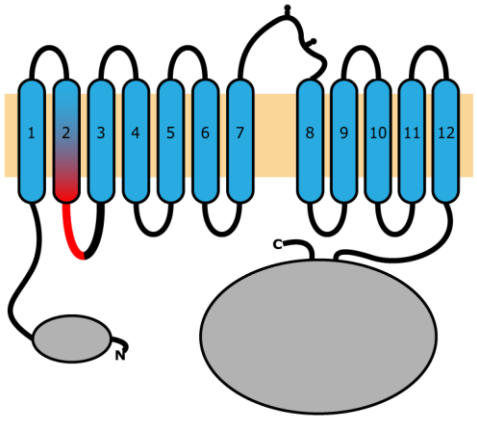


Fig 1-10 Simplified Sheme of the Topology of NKCCs
NKCCs are proposed to consist of 12 transmembrane helices. The small N-terminal domain and the C-terminal domain are both located in the cytoplasm. Two glycosylation sites were identified between transmembrane segment 7 and 8. In NKCC2 an exclusive exon cassette starting in the middle of transmembrane helix 2 to the middle of the connecting intracellular loop defines the splice variants A, B and F (red).

domain that constitutes nearly half of the protein (413-480 residues). The transmembrane domain is predicted to harbor 12 transmembrane spanning segments (Fig 1-10). For the three different groups of CCCs distinct glycosylation sites have been identified. NKCCs and NCC have two glycosylation sites in the loop between helix 7 and 8 and KCCs four sites between helix 5 and 6.

Sequence Comparisons between all members of the CCC superfamily reveal valuable information about the homologies within the family and the conservation of the three topological domains (N-terminus, transmembrane domain and C-terminal domain). The pair-wise sequence comparisons of the full length proteins gives rise to the already mentioned clustering into two subfamilies of Na^+- and K^+-driven family members (see above 1.2.1). While most of the putative helices align well between these two groups, clear differences in loop lengths are apparent. Particularly the respective loops harbouring the glycosylation sites are longer in KCCs. Overall the sequence conservation is highest for the membrane domain and lower at the C-terminal domain. Among the three parts of the protein the N-terminal tail is least conserved. The membrane domain contains the specific binding sites for the substrates and it catalyzes the transport and thus shows the highest sequence conservation (Fig 1-6B). The C-terminal domain, in contrast, has important regulatory functions for activation and deactivation by phosphorylation or potential ligand binding and serves as interaction platform with other proteins (Moore-Hoon and Turner 2000; Mercado, Broumand et al. 2006). Additionally the C-terminal domain was suggested to play a role in the subunit recognition and thus to prevent hetero-oligomerization (Parvin, Gerelsaikhan et al. 2007). This demand for selectivity could explain the significant lower conservation. The amino terminus that is the most variable segment in length has a reasonable identity among KCCs (~45 %) but is not conserved between the subfamilies (<30 %). Nevertheless, it was shown to have important regulatory functions via phosphorylation (Gimenez and Forbush 2005; Anselmo, Earnest et al. 2006).

The best characterized members of the CCC family are the sodium dependent transporters NCC and NKCC1/2. They were the first identified electroneutral cotransporters, are abundant among the animal kingdom and have a distinct expression pattern in mammals. NKCCs are particularly suited for functional investigations because they exist in different isoforms and have splice variants with distinguishable properties.

NKCC2 is of specific interest because three splice variants exist (NKCC2A, NKCC2B, NKCC2F) that each contains a variation of 31 residues (Fig 1-11C) (Gamba 2005). This variation concerns a region in the C-terminal half of transmembrane helix 2 and the first part of the consecutive intracellular loop and is responsible for differences in the ion transport characteristics of the splice variants. Additionally, a long (NKCC2-l), (Fig 1-11A) and short version (NKCC2-s), (Fig 1-11B) exist that differ in the length of their C-terminus (329 residues to 55 residues, respectively). NKCC2A, B, F are distributed along TALH according to their different ion affinities and transport rates. In

the outer medulla where the primary urine still has a high salt concentration the isoforms F with the lowest affinity and highest transport rate is expressed. Isoform A, which is spread all along TALH, shows high transport rates with moderate affinity. Isoform B possessing the highest affinity for the transported ions and is solely expressed in the cortex, where the ion concentration in the tubular fluid has been greatly reduced. On a sequence level, these changes in affinity (factor 3 to 7 for Na^+, 2 to 3 for K^+, 3 to 12 for Cl^- for mouse and rabbit, respectively) are accomplished by minor changes of only three residues (Fig 1-11C).

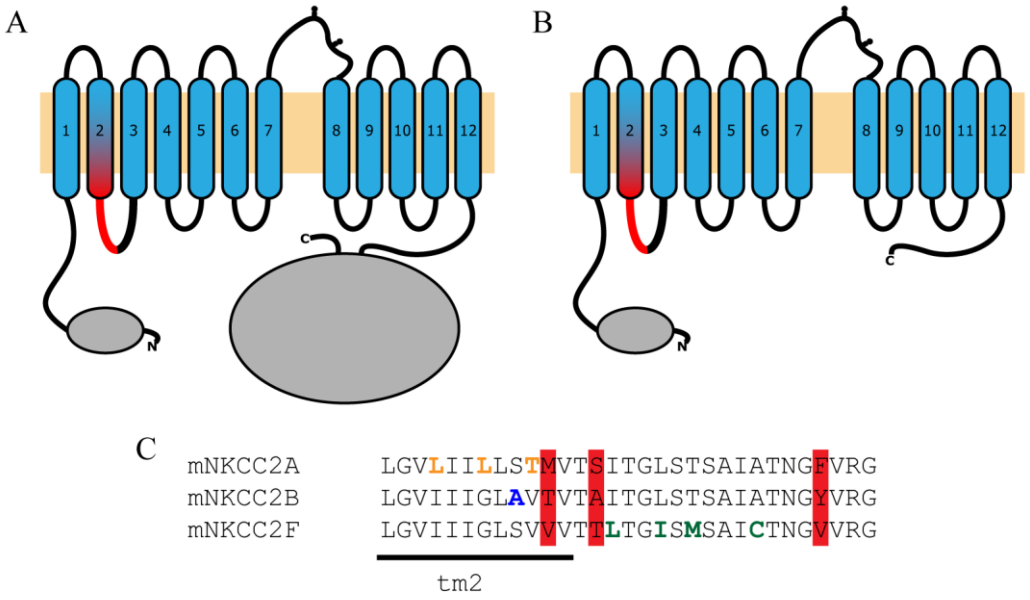


Fig 1-11 Splice Variants of NKCC2
NKCCs are proposed to consist of 12 transmembrane helices. The small N-terminal domain and the C-terminal domain are both located in the cytoplasm. Two glycosylation sites were identified between transmembrane segment 7 and 8. Due to splice events a long NKCC2-l **(A)** and a short NKCC2-s exist **(B)**, which differ in the length of their C-terminus (329 residues in the long and 55 residues in the short isoforms). **(C)** Within the long and the short version of NKCC2 a region starting in the middle of transmembrane helix 2 (tm2) to the middle of the following intracellular loop defines the splice variants A, B and F (illustrated for mouse NKCC, mNKCC). Each splice variant is distinguished by three residues that are changed in all exons (red boxes) (orange-A, blue-B, green-F).

The physiological role of NKCC2-s is still poorly understood. Functional experiments in *Xenopus laevis* oocytes suggest two different roles: as regulatory molecule and as cotransporter. As regulatory molecule, NKCC2-s has a negative effect on NKCC2-l function that can be abrogated by cAMP. The presumed mechanism, as shown by fluorescence analysis, is the prevention of NKCC2-l to reach the plasma membrane. The observed activation of NKCC2-l by cAMP can be interpreted as an effect of cAMP producing hormones like vasopressin via their G protein-coupled receptors (GPCRs). NKCC2-s also appears to function as K^+-independent Na^+/Cl^- cotransporter as shown in *X. laevis* oocytes with murine NKCC2-s. This behavior was

observed in hypotonic solution and was abolished by exposure to an isotonic environment or stimulation of cAMP production via vasopressin. In these cases, NKCC2-l was the major expressed protein and the apical transport switched to $Na^+/K^+/Cl^-$ cotransport.

The human and shark orthologues of NKCC1 share high identity (74 %) but have different affinities for the transported ions (six times higher for hNKCC1) and for bumetanide inhibition. Different chimaeras of the two organisms were analyzed for their transport characteristics in transfected HEK-293 cells (Isenring and Forbush 2001). The hybrid constructs were compared with the wild type to identify transmembrane segments responsible for ion and inhibitor binding. The transmembrane region (Tm) 2 plays an important role in Na^+ and K^+ transport, while Tm 4 was shown to influence the K^+ and Cl^- transport characteristics. Tm 7 is crucial for the transport of all three ions: Na^+, K^+, and Cl^-. Bumetanide binding, in contrast, could not be assigned to any specific region (Fig 1-12).

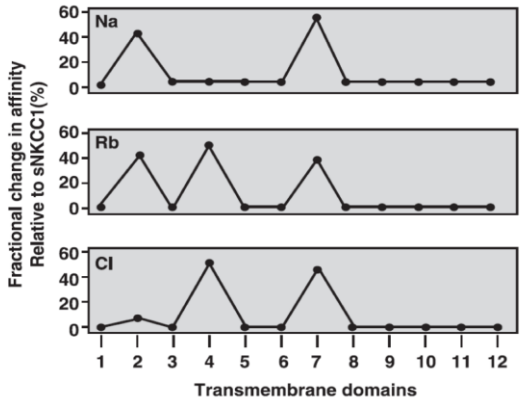


Fig 1-12 Distribution of Ion Specificity on Transmembrane Segments of NKCC1
Relative change in the transport characteristics of hNKCC1 - shark NKCC1 (sNKCC1) chimeras. The affinity change of the chimera (compared to sNKCC1) for the transported ions (Na^+, Rb^+, Cl^-) are plotted against the predicted transmembrane domains. Rb^+ is used as replacement for K^+. (Isenring and Forbush 2001)

The high presence identity between the orthologues, with variation in single residues allowed the identification of few amino acids within the segments that account for the observed changes. During these experiments also the influence of the cytosolic N- and C-terminal domains on transport was also investigated. In these investigations no changes in transport properties and inhibition characteristics upon altering the cytosolic domains could be detected. The ion and inhibitor affinities are thus solely determined by the transmembrane domain. Similar experiments were performed using the rat and winter flounder orthologues of NCC which have led to the same conclusions.

1.2.3.1 Prokaryotic Homologues as Model Proteins

High throughput sequencing efforts of organisms provide a growing repertoire of prokaryotic genomes. The identification of prokaryotic homologues of eukaryotic transport proteins has accelerated the structural investigation of membrane proteins for several reasons. First, because prokaryotic expression systems, particularly *E. coli*, have been much more investigated and developed than eukaryotic expression systems. Second, because prokaryotic proteins are easier to overexpress and purify in large quantities. In contrast to their eukaryotic counterparts these proteins mostly lack extensive, unstructured loops, posttranslational modification like glycosylation and are usually not phosphorylated. A particularly promising class of proteins derives from extremophile sources like thermophiles, because they are believed to harbor very stable proteins important for solubilization, purification and crystallization. Working with various prokaryotic homologues thus offers the possibility to test related proteins that share the same structure and function but have remarkably different properties. Prokaryotic homologues have traditionally been called evolutionary related and to belong to the same family if they have a sequence identity >30 % in a pair-wise alignment (Granseth, SeppÃ¤lÃ¤ et al. 2007). Identity above this margin is a clear evidence for a functional and structural relationship. However, during the last years this threshold has been found to be significantly lower. Functional experiments that have led to the identification of essential residues and information of architectural features of the proteins by biochemistry have allowed reducing the cut-off for identical residues. Various structures of prokaryotic membrane protein homologues have previously been determined with lower conservations to their eukaryotic counterparts that clearly belong to the same family. In the case of ClC chloride channels and transporters the sequence identity between the human family members is already relatively low with 21.9 % (hClC-0 to hCLC-5). The prokaryotic homologue of *E. coli*, EcClC, has a sequence identity to hClC-5 of 16.6 %. Nevertheless the sequence conservation is high in regions of functional importance. The structure of EcClC allows the localization of conserved residues of which some form the ion translocation pore with specific ion binding sites.

For CCCs a number of prokaryotic homologues could be identified. They share a unusually high degree of sequence identity (above 30 %) to their human counterpart NKCC2 (Fig 1-13). The conservation is especially high in the transmembrane domain (residue 178 to 813 for NKCC2). Unlike their mammalian counterparts, the prokaryotic transporters lack extensive loops between defined structural elements (loop residue 345-356 and loop residue 439-475 for NKCC2) and for prokaryotic CCC homologues the N terminal domain (residue 1-174) (Fig S-1).

The insights gained from the structures of prokaryotic homologues have guided the interpretation of functional experiments on eukaryotic family members. Particularly the basic transport mechanisms and substrate specificity are usually conserved between the pro- and eukaryotic species barrier. The prokaryotic structures of membrane proteins thus also provide a structural framework for follow-up experiments on the eukaryotic

counterpart. While general mechanisms are conserved, certain regulatory and pharmacological properties often differ between pro- and eukaryotic transporters. The cellular regulation of specific family members, the binding of drugs and the role of unstructured regions will ultimately require the structural investigation of eukaryotic CCCs.

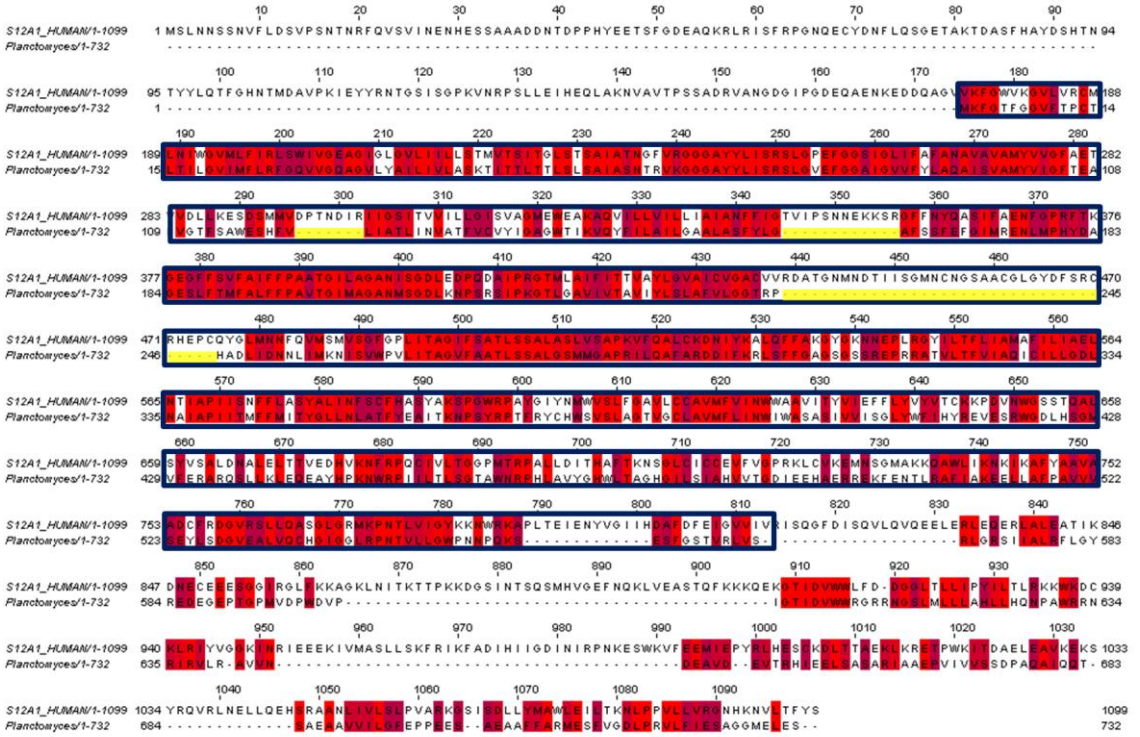


Fig 1-13 Sequence Alignment Between Human and Prokaryotic CCCs
Human NKCC2 (S12A1) was aligned with a prokaryotic homologue from *Plactomyces maris* (37 % identity). The numbering (above) corresponds to the residue number of S12A1. The coloring is according to conservation (red=identical residues, ruby=very homologues residues). The region of highest conservation is located within the transmembrane domain (residue 174 to 813 for S12A1, blue box). Insertions in presumed loop region within the eukaryotic transporter are colored in yellow.

1.2.4 Crystallization and Structure Solution of Membrane Proteins

The knowledge of detailed spatial arrangements of amino acids in the structure of a protein is crucial for the in depth analysis of its function. Two methods are widely used to extract this information: X-ray diffraction of protein crystals and nuclear magnetic resonance (NMR) in solution. NMR offers the possibility to extract information on dynamic behavior of the protein and it does not require crystals but it is limited by the size of the biopolymer (around 30 kD). Crystallography in contrast freezes one conformation of the protein in a crystalline array and thus allows the detailed structure determination of proteins and complexes even exceeding 1 Mega-Dalton.

The structural analysis of proteins demands milligram amounts of pure protein. This requirement is a major hurdle towards structure determination. The expression, purification and crystallization of membrane proteins are particularly difficult for a number of reasons. Due to their localization in membranes these proteins expose a hydrophobic surface and can thus only be solubilized in the presence of detergents which results in poor stability and increased flexibility in solution.

Membrane proteins are usually either purified from natural sources or heterologously overexpressed in different expression systems. Eukaryotic proteins of known structure have been mainly purified from tissues where they are naturally enriched. Rhodopsin was isolated from retina, Na^+/K^+ ATPase from kidney, aquaporins from blood, and cytochrome oxidase from heart. Many prokaryotic membrane proteins and recently also eukaryotic membrane proteins have been successfully overexpressed in different pro- and eukaryotic expression systems, prokaryotic proteins in *Escherichia coli* and *Lactococcus lactis* and eukaryotic proteins in the yeasts *Saccharomyces cerevisiae* and *Pichia pastoris*, insect cells and in mammalian cell lines. The expression systems are sorted by increasing complexity with respect to the translation machinery, posttranslational modifications and lipid composition of the membrane. These factors have become important for functional expression of complex eukaryotic proteins.

Membrane proteins are extracted from membranes by detergents. Detergents are amphiphilic molecules composed of a polar or charged head group and a single hydrophobic tail. They mimic the amphiphilic nature of the lipid membrane and are thus suited to cover the hydrophobic belt of the membrane protein. Due to their cone-like shape, detergents spontaneously form micelles in solution if the concentration is above the critical micelle concentration (CMC). These micelles surround the hydrophobic

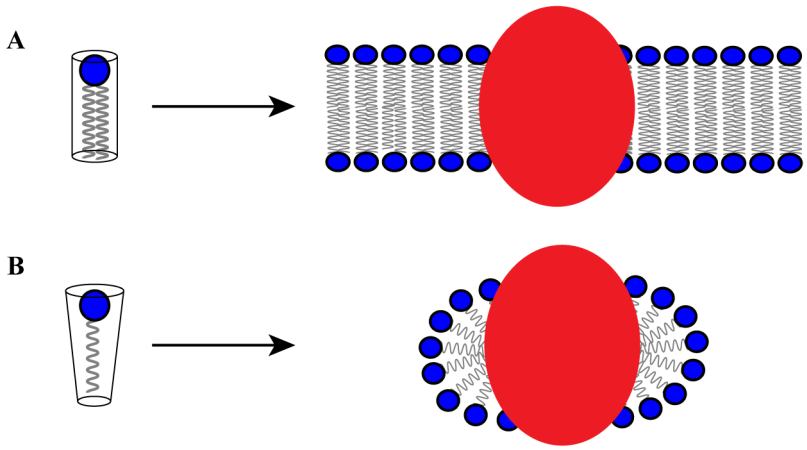


Fig 1-14 Characteristics of Amphiphiles
Amphiphiles are composed of a hydrophilic head group (blue) and a hydrophobic tail (grey). Depending on the volume ratio of head group to tail of a monomer they occupy a cylindric shape (**A**; lipids) that results in a 2D layer or a cone shape (**B**; detergent) creating a globular micelle. The micelle formation starts at concentrations around the critical micelle concentration (CMC).

outside of the membrane protein and thus convert the previously insoluble protein into a soluble protein detergent complex (Fig 1-14B). The use of the appropriate detergent is critical for the purification process. They are chosen based on their ability to extract the protein in an active, homogenous and stable state. The solubilized detergent-protein complex is usually crystallized by standard methods.

The crystallization process is little understood and is mainly a trial and error procedure. In this study initial crystallization conditions are obtained using robot screening of thousands of conditions in a vapor diffusion experiment. Promising conditions are subsequently refined by hand to obtain useful diffraction. In an alternative approach crystallization of membrane protein can also be achieved using a lipid environment like lipidic cubic phases, lipid sponge phases and bicelles instead of detergent micelles (Johansson, Wöhri et al. 2009).

An important step towards the identification of well diffracting crystals is the modification of the protein construct and the modification of purification protocols to improve identified crystal forms. Such modification include the addition of lipids to the purification buffer, recloning of the protein to eliminate floppy parts, the increase of the thermo stability by introduction of single point mutations and the monitoring of the state of the protein during expression and purification by attachment of green fluorescent protein (GFP) (Geertsma, Groeneveld et al. 2008) (Kawate and Gouaux 2006).

1.2.5 Motivation

While transport properties of this important class of membrane proteins have been investigated during the last years, a thorough understanding of the structure-function relationship is still pending due to a lack of structural insights into this family. Until now the only available structural information comes from the prediction of membrane topology and secondary structure elements. The goal of this thesis is to establish the overexpression and purification of a prokaryotic CCC transporter as model system for structural and functional investigations to reveal first insight into the structural organization of the family.

2 Results

2.1 Overview

The first stage in the process to obtain structural and functional information about a membrane protein is the production of functional and stable protein in large amounts. This process is accompanied by various challenges including mRNA stability, efficiency of translation and numerous posttranslational modifications of the protein. In later stages the homogeneity and stability of the protein in detergent solution poses challenges. To overcome these bottlenecks different strategies were combined (Fig 2-1).

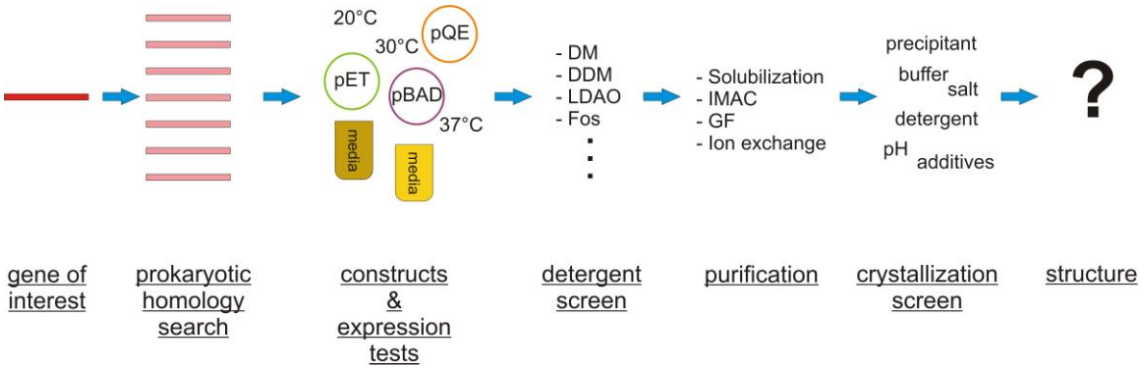


Fig 2-1 Pipeline Towards the Structure Solution of a Membrane Protein

First, I have used established prokaryotic expression systems for overexpression of the target membrane proteins such as *E. coli* and *L. lactis*. Since different homologues of the same family have unique properties, I have included family members from diverse organisms that were identified by homology search within the microbial genomes. I have limited my efforts towards microbial proteins since these proteins do not contain certain characteristics of eukaryotic proteins that might prevent crystallization such as flexible loops, glycosylation sites and extra unstructured domains at the termini. The set of proteins chosen for overexpression and purification includes CCCs from prokaryotic extremophiles that are believed to contain very stable proteins that survive under extreme conditions, features that are believed to be beneficial for purification and crystallization.

With the choice of *E. coli* and *L. lactis* as expression systems I cover Gram-negative and Gram-positive expression hosts. For the overexpression in *E. coli* I have used the strong lactose promoter system that controls expression of the T7 RNA-polymerase and the tunable arabinose promoter (pBAD) that directly allows the controlled expression of

the target protein. *E. coli* allows investigating the protein expression in a broad range of temperature, media and inducer concentrations, which might be beneficial for the

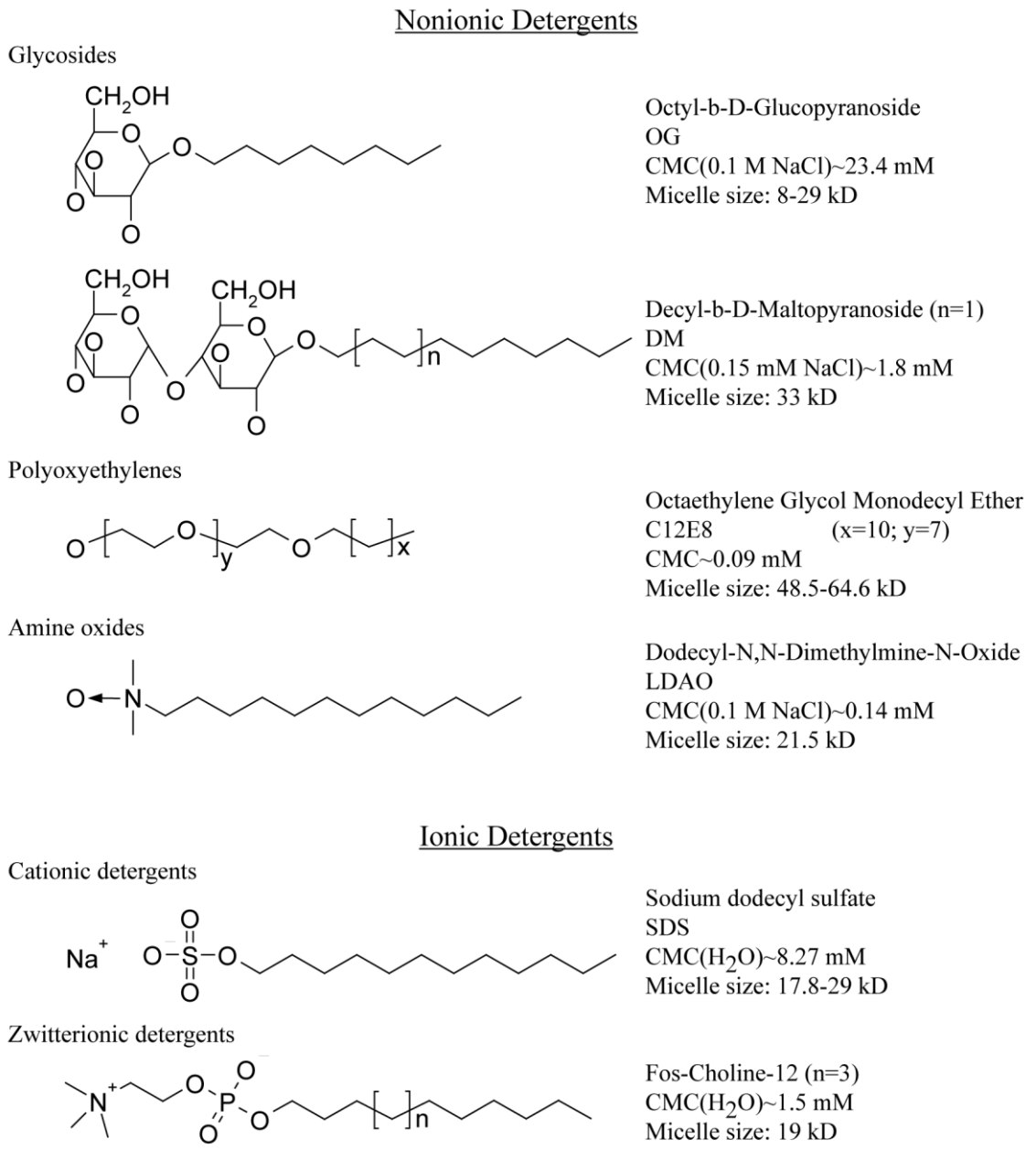


Fig 2-2 Detergents
Structures and properties of commonly used detergents in membrane protein research.

production of functional protein.

After overproduction of the membrane protein it must be extracted from the membrane to solution with detergent that mimics the natural environment of the protein in the hydrophobic membrane of the cell. The choice of the right detergent is the next challenge in the pipeline. Two constrains have to be fulfilled: The detergent must extract the protein from its membrane environment and it must stabilize the protein in

solution during the course of the experiment. Detergents differ in size and polarity of the head group and in the length of their hydrophobis tail (Fig 2-2). Nonionic detergents are mostly used for purification and crystallization, especially glycosides such as glucosides and maltosides that lead to the majority of successful structure solution of membrane proteins.

Since the transporters are expressed as fusion proteins to a polyhistidine tag (His-tag), the following purification steps include immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) to produce clean and homogenous protein. SEC is additionally used to test the dispersity of the protein.

All described parameters are crucial for the overexpression, purification and crystallization of prokaryotic members of the CCC family that will be discussed in detail in the following paragraphs.

2.2 Homology Search

During the last decade the number of sequenced microorganisms increased from a few to over thousand of genomes. This increase in genomic information offered the opportunity for extensive homology search for prokaryotic CCCs. The protein sequence of human NKCC2 was chosen as query for blast searches (blastp). The Blossum62 matrix was used for the scoring of the alignments and all available microbial genomes at a time were chosen as query database. The first search at the beginning of this work pulled out 3 homologous proteins from the bacteria *Cadidatus Protochlamydia amoebophila* (CpCCC) and *Synechococcus sp.* (SynCCC), and the archaea *Methanosarcina acetivorans C2A* (MaCCC) (Tab 2-1). All three proteins share around 30 % identical residues. Additionally four proteins were identified with lower homology (<25 %) (NAMES OF THE PROTEiNS). Due to the increasing size of the genome database the latest expression and purification trials included nearly 20 homologues.

Strain	Strain code	Identity [%]	Similarity [%]	Abbr.	Length [aa]	Gene	M_w [kD]
Homo sapiens	-	100	-	NKCC2	1099	SLC12A1	121.5
Planctomyces maris	DSM 8797	36	55	PmaCCC	732	PM8797T_10374	80
Syntrophobacter fumaroxidans MPBO	DSM10017	34	52	SfCCC	733	Sfum_3620	80.5
Salinibacter ruber	DSM13855	30	51	SrCCC	754	SRU_2259	81
Cyanothece sp.	CCY0110	31	52	CyaCCC	746	EAZ91226	81.5
Crocosphaera watsonii	WH8501	31	52	CwCCC	744	EAM51764	81
Psychromonas ingrahamii 37	DSM17664	28	50	PiCCC	847	Ping_1733	93
Trichodesmium erythraeum	IMS101	31	52	TeCCC	744	Tery_1903	81
Lyngbya sp.	PCC8106	30	52	LyCCC	747	EAM34907	81
Synechococcus sp.	PCC7002	30	50	SynCCC	732	SYNPCC7002_G0123	80
Cadidatus Protochlamydia amoebophila	UWE25	29	48	CpCCC	764	pc0822	84.5
Methanosarcina acetivorans C2A	DSM2834	31	47	MaCCC	758	MA4506	83
Salinibacter ruber	DSM13855	29	48	Sr2CCC	815	SRU_2257	87

Tab 2-1 Prokaryotic Homologues
List of 12 homologues with the best alignment properties (05-2009). The homologues were identified using blastp with Blossum62 matrix searching with NKCC2 (SLC12A1). Identity and similarity were determined in a pair-wise alignment with NKCC2.

The identified proteins were examined with respect to the homology to human NKCC2, taking their homology scores, identical and similar residues and the distribution/clustering of homologous regions as criteria. The overall very high amount of conservation is striking. Among prokaryotic membrane proteins often the sequence identity with their eukaryotic counterpart does not exceed 15 %. With a sequence identity of around 30 % and a sequence homology approaching 80 % in a pair-wise alignment to NKCC2 the conservation between pro- and eukaryotic CCCs is unusually high which makes it safe to propose that the prokaryotic proteins belong to the same

family of proteins and share the overall fold and transport mechanism. The identified homologous proteins are distributed among archaea (MaCCC) and bacteria (*e.g.* PmaCCC, SfCCC, SrCCC). Cyanobacteria harbour the largest fraction of close homologues like CyaCCC and CwCCC. SfCCC and MaCCC are both derived from extremophiles which makes them interesting targets for structural investigations. The overall length of the bacterial transporters is around 750 amino acids which is about 350 residues smaller than human NKCC2. This reduced length is a result of missing residues at the N-terminus of the protein as well as in loop regions between defined secondary structure elements.

Upon colder inspection of the alignment interesting features become apparent which

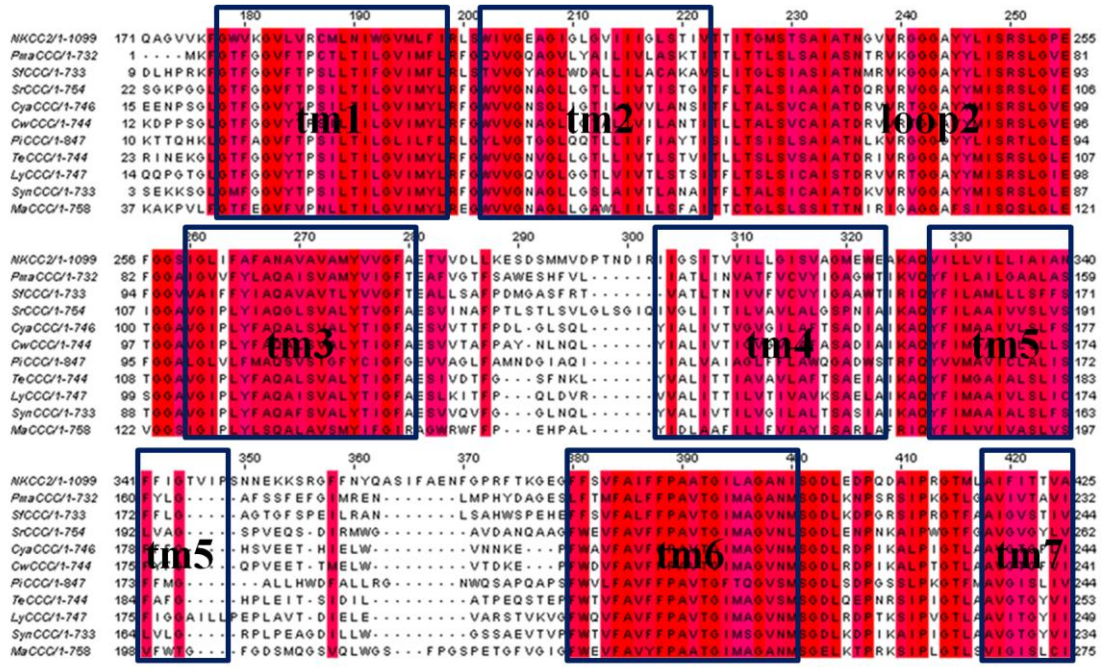


Fig 2-3 Sequence Alignment of Selected Prokaryotic Homologues
Human NKCC2 is aligned with the prokaryotic homologues PmaCCC, SfCCC, SrCCC, CyaCCC, CwCCC, PiCCC, TeCCC, LyCCC, SynCCC and MaCCC (from top to bottom). The numbering (above) corresponds to the residue number of human NKCC2. The coloring is according to conservation (red=identical residues, ruby=very homologues residues). The structural elements are labeled according to the prediction for NKCC2. The region of highest conservation is located within the transmembrane domains tm1-tm7 (blue box) and the first intracellular loop (loop2). Insertions in presumed loop regions within the transporters are visible between tm3 and tm4 as well as between tm5 and tm6.

underline the conservation of functionally important regions (Fig 2-3). Following the cytoplasmatic N-terminal domain of NKCC2 12 highly conserved stretches of 18 to 22 amino acids are found in the proposed transmembrane part of the protein. The predicted transmembrane spanning regions are interrupted by connecting loops that, in general, show lower conservation. Only the first intercellular loop, between transmembrane segment 2 and 3, is highly conserved. This loop is predicted to be essential for the ion

transport by CCCs. In contrast to the large conservation in the membrane domain, the conservation in the C-terminal domain is moderate with 10-13 % of identical residues.

2.3 Prokaryotic Full-Length CCC

2.3.1 Cloning

Samples of the selected microbes were purchased from various collections of prokaryotic organisms such as DSMZ, ATCC and PCC, a collection of cyanobacterial strains of the Pasteur Institute as dried cell pellets or growing cultures. The genomic DNA was subsequently isolated by standard procedures resulting in sufficient starting material for amplification. The genes of interests were amplified by the polymerase chain reaction (PCR), using primers that contain restriction sites for the subsequent cloning in expression vectors. The PCR protocol was adapted to take the differences in the nucleotide content into account. Besides variation of the PCR buffer composition the choice of the DNA polymerases (*e.g.* Vent®, Pfu®, Phusion® or mixures of polymerases as present in the Master Mix®) had the biggest influence on successful amplification.

In the described work three different cloning strategies were pursued. The different gene products were cloned into *E.coli* expression vectors using either traditional Class II restriction enzymes, with overlapping, palindromic recognition and cutting sequences or fragment exchange cloning (FX cloning). To obtain expression constructs for *L. lactis*, vector backbone exchange (VBex) allowed the transfer of the gene of interest into a *L. lactis* expression vector.

For cloning with ClassII restriction enzymes, a modified pET28b+ vector (Novagen) containing a 3C protease site sequence followed by a His-tag sequence at the 3’ end of the gene was used (Fig 2-5A). The preferred 5' restriction sites were Nco1 and Nde1, which already contain a start codon, and 3' restriction sites like EcoR1 or Xho1. With this strategy no additional residues were introduced at the N-terminus of the produced protein. One drawback of this strategy is the use of two restriction enzymes for directional cloning of the amplified gene into the vector and the abundance of the respective restriction sequence in the gene of interests which prevents the use of specific enzymes in certain cases.

To circumvent the problem of internal cleavage sites and to facilitate high throughput cloning efforts a different strategy was followed. FX cloning only uses one enzyme for directional cloning. As for the Gateway cloning system, the genes of interest and the target vector are mixed in one tube to which the enzyme is added. The enzyme catalyzing the reaction is very specific resulting in a low frequency of site in the genes of interest that could interfere with the procedure.

Another beneficial feature for high throughput cloning is the integration of the selection marker Ccdb in the original plasmid (Fig 2-4B). Ccdb is a cytotoxic protein that inactivates DNA gyrase and that is integrated into the plasmid under the control of

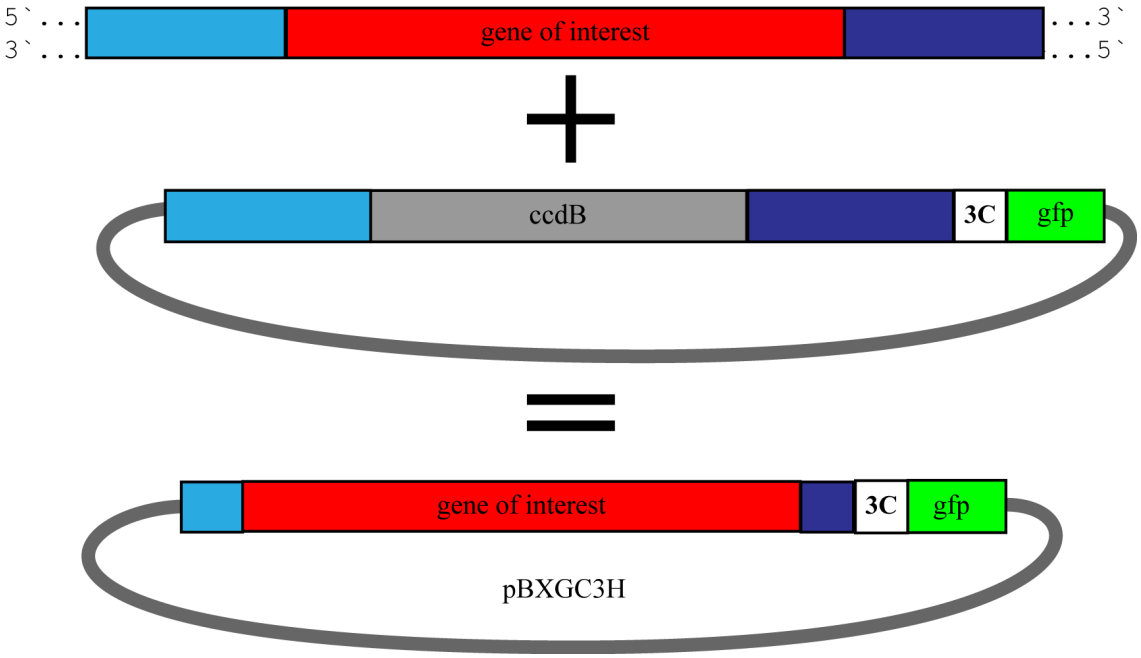


Fig 2-4 FX Cloning
Fragment exchange (FX) cloning requires the addition of small DNA sequences on both sides to the gene of interest (blue). They differ in sequence to guarantee directional cloning. The target vector contains the CcdB marker (grey) flanked by the corresponding small DNA recognition sequences. After the cloning reaction 3bp are left on the plasmid introducing an additional amino acid on both sides (5` serin, 3` alanine). CcdB (grey) is a toxic protein to *E. coli* that prevents growing of false negative clones.

a consecutive promoter (Bernard, Gabant et al. 1994). This cassette is eliminated after successful ligation with the insert. Thus relegations and uncut vectors are eliminated since bacteria containing such vectors fail to grow. N-terminal (pBXN3H) and C-terminal (pBXGC3H) His_{10}-tag fusions were constructed. This cloning method introduces an additional amino acid at the N-terminus of the protein (Fig 2-5B).

L. lactis expression plasmids must be constructed differently with VBex since cloning in *L. lactis* is very laborious (Geertsma and Poolman 2007). The major manipulations are carried out in *E. coli* with an appropriate shuttle-vector that includes an origin of replication for *E. coli* (pREX). The first step is similar to the previously described FX cloning. In the second step the purified pREX plasmids of positive clones were mixed with a plasmid containing the essential replication elements for *L. lactis* (pERL). In the following reaction the vector-backbone exchange takes place. The mechanism is similar as for FX cloning but uses another enzyme Sfi1. Sfi1 has a freely designable cutting sequence of 5 bp that is surrounded by a 4 bp recognition sequence

on each site. The two plasmids pREX and pERL have both two Sfi1 sites for directional cloning. During digestion of both vectors in the same tube, the religation and exchange of the vector fragments take place. A fraction of the generated plasmids combine the *L. lactis* origin of replication and selection marker with the gene of interest to create the resulting plasmid pNZ. The mixture is transformed into *L. lactis* by electroporation and cells that carry the correct plasmid are selected for by addition of antibiotics. In some cases of the constructs obtained by FX cloning a C-terminal fusion of green fluorescent protein (GFP) to the protein of interest was generated. In these fusion proteins the GFP portion can be used as folding indicator, for detection and quantification of the protein (see Fig 2-6).

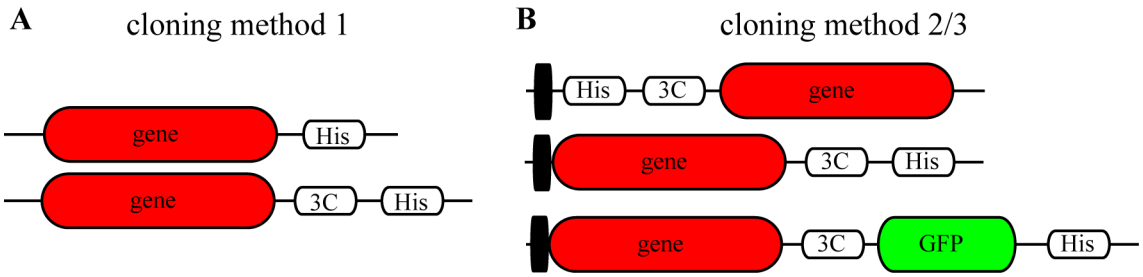


Fig 2-5 Expression Constructs for Prokaryotic Homologues
The genes of interest (red) were cloned using different methods and constructs. **A** Cloning by ClassII restriction enzymes (cloning method 1) resulted in constructs with a C-terminal His-tag (top), or a C-terminal His-tag separated from the protein by a 3C protease site (bottom). **B** Using FX cloning and FX/VBex cloning (cloning method 2 and 3) fusion proteins containing N- and C-terminal His-tags separated from the protein by 3C protease sites were generated (top and center). In certain constructs, GFP (green) was introduced at the C-terminus of the protein (bottom). FX cloning methods introduce an additional amino acid on the N-terminus of the protein (black box).

2.3.2 Expression Tests

The identification of proteins with high expression levels and suitable biochemical properties is one of the critical steps towards structure determination. For that reason it was important to be able to broadly analyze the expression behavior of different protein constructs parallel and in low volumes. The expression tests were carried out with *E. coli* and *L. lactis* as expression hosts. The expression volume ranged from 1 ml in 96 well plates to 5 ml in culture tubes. For comparison similar quantities of cells or isolated membranes were analyzed to correlate the amount of overexpressed protein for different protein constructs. In general 20 µg of total protein were analyzed corresponding to 2 µg of total membrane protein. The two techniques for examining the expression level were anti His-tag Western blot analysis and the analysis of GFP fluorescence. The two methods are complementary and reveal different properties of the expressed proteins.

For anti His-tag Western blot analysis a whole-cell extract is separated by SDS-PAGE and the gel is blotted on a PVDF membrane following standard Western blot protocols. With this method all proteins that contain a N- or C-terminal His-tag fusion

can be detected. A disadvantage of the method is the difficulty in the comparison of signals between different gels and the quantification of the signal for a correlation with the total amount of expressed protein. Additionally, no conclusions on the quality of the expressed protein can be drawn since misfolded or aggregated proteins are equally well recognized. To circumvent these limitations a sedimentation assay was included to remove aggregated protein and in that way to obtain information on the quality of the expressed material. After expression the cells were broken and membrane proteins were extracted by addition of detergent (typically 2 % DDM [w/v]). After centrifugation of the sample at high velocity (100000 x g) the amount of His-tagged protein in the supernatant was compared with the same protein before centrifugation. While well-folded protein stays in the supernatant after high spin centrifugation, aggregated protein sediments faster and is usually found in the pellet.

The sedimentation assay can be circumvented with the second protocol that makes use of GFP fluorescence (Geertsma, Groeneveld et al. 2008). For this assay fusion proteins that have GFP attached to the C-terminus of the protein of interest are used. The fluorescence properties of cell extracts that were separated on a SDS-PAGE gel were analyzed. Since GFP does not unfold under SDS-PAGE conditions, fusion proteins do in this case still show fluorescence. This method allows the quantification of the expressed protein levels and is thus suited for a direct comparison between different expression conditions and different homologues. Additionally, GFP that is fused to the C-terminus of the protein of interest can be used as folding indicator to detect the quality of the preceding protein. During translation GFP only folds and shows fluorescence if the preceding target protein is properly folded. If the protein, however, is misfolded and thus expressed in an aggregated form also GFP is misfolded and will not show any fluorescence. Well-folded protein in contrast allows proper folding of GFP and thus allows the quantification of the expression level by recording of its fluorescence. This method was applied for the measurement of in-cell GFP fluorescence in a 96 well plate format and for the measurement of on-gel fluorescence after SDS-PAGE. This workflow opens the possibility for high through-put analysis of well-folded protein expression.

The combination of in-gel GFP fluorescence analysis and anti His-tag Western blot analysis of the same SDS-PAGE gel provides complementary information (Fig 2-6). Folded GFP connected to a SDS-denatured target-protein has a different migration behavior in SDS-PAGE than unfolded GFP fused to its target protein. The compact, folded GFP migrates faster than the misfolded protein and is thus separated from not properly folded material. After detection of the in-gel fluorescence signal the gel was further analyzed with anti His-tag Western blot. While in-gel fluorescence only shows the folded fraction of the GFP-tagged protein the Western blot detects both populations

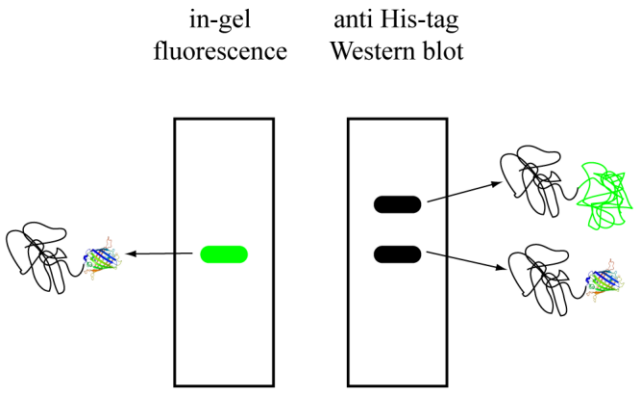


Fig 2-6 GFP as Expression and Folding Indicator
Green fluorescent protein (GFP) can be used as folding indicator. A C-terminal GFP fusion to the protein is in the same state as the preceding protein during expression. If the preceding protein is folded, GFP is also folded and shows fluorescence. In contrast, if the preceding protein is aggregated, GFP is misfolded and lacks fluorescence. Since GFP keeps its conformation during SDS-PAGE analysis, it can be used to directly trace the folded fraction of the target protein within the gel (left). Since folded and unfolded GFP fused to SDS denature protein have different migration behaviors, they can be separated and thus used to analyze the ratio of folded (right, lower band) to aggregated protein (right, upper band) expression.

of the protein and thus allows the discrimination between well-folded protein (lower band) not folded protein (upper band).

2.3.2.1 Overexpression of CCCs in *E. coli* Under the Control of the P_{BAD} Promoter

The use of the araBAD promoter offers an expression strategy that allows the tight regulation with a dose-dependent induction that combines gentle expression conditions with high expression yields. The inducer in this case is arabinose. In arabinose-free medium the AraC dimer binds two distant parts of the araBAD-operon and suppresses transcription through DNA loop formation. Upon arabinose binding the AraC dimer rearranges and binds to near parts of the araBAD-operon thus releases the loop and stimulates expression of the target gene. For this induction system *E. coli* strains that are incapable of metabolizing arabinose due to a disrupted *ara* gene such as TOP10 and MC1061 are available. MC1061 was the main cell line used for the described expression tests. The first constructs tested were prokaryotic homologues with C-

terminal GFP fusion followed by a His_{10}-tag. The expression medium in these studies was Terrific broth (TB) with variation of arabinose concentration and temperature.

The expression was analyzed by comparing in-gel GFP fluorescence and anti His-tag Western blot luminescence to judge the quality and quantity of expressed protein in dependence of expression conditions.

First tests were performed using transformed *E. coli* MC1061 that were grown at 37 °C to an $OD_{600\ nm}$=0.4, cooled down to 25 °C and induced with 10^{-1} to10^{-5} % (w/v) arabinose. After overnight growth at 25 °C the cells were harvested and equal amount of cells were analyzed. The expression of twelve different bacterial homologues of CCCs that contain C-terminal GFP fusions was investigated. The in-gel fluorescence showed that five out of twelve homologues, MaCCC, CpCCC, PmaCCC, PiCCC and SynCCC express well-folded protein with only minor degradation levels while the expression of the remaining seven homologues (CwCCC, LyCCC, CyaCCC, SrCCC, Sr2CCC, SpCCC, and SfCCC) was too low to proceed (Fig 2-7). The anti His-tag immunostaining revealed differences between the homologues. While the ratio of folded to misfolded protein is high for MaCCC, PmaCCC and PiCCC and somewhat for

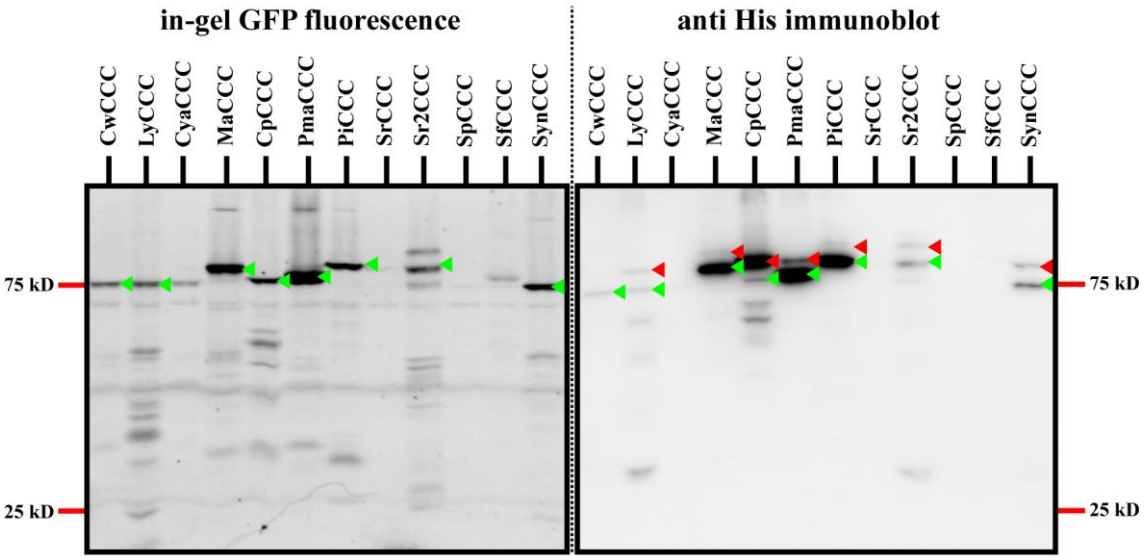


Fig 2-7 Analysis of Expression in *E. coli* Using GFP Folding Assay
Homologues expressing in *E. coli* were selected for comparative analysis using the GFP folding assay. Expression was evaluated by in-gel GFP fluorescence (left panel) and anti His immunoblotting (right panel). Whole cell extracts corresponding to 20 µg total protein were separated via SDS-PAGE after cell lysis. Green triangles indicate well-folded fluorescent protein whereas red triangles mark misfolded, non fluorescent protein. Both analyses were performed using the same SDS-PAGE gel.

SynCCC. CpCCC showed a major fraction of misfolded protein that manifests itself as upper band on the Western blot (Fig 2-7). This homologue was nevertheless included for continuing investigations, since the amount of expressed protein was comparably high and since a variation of expression conditions might still improve the result.

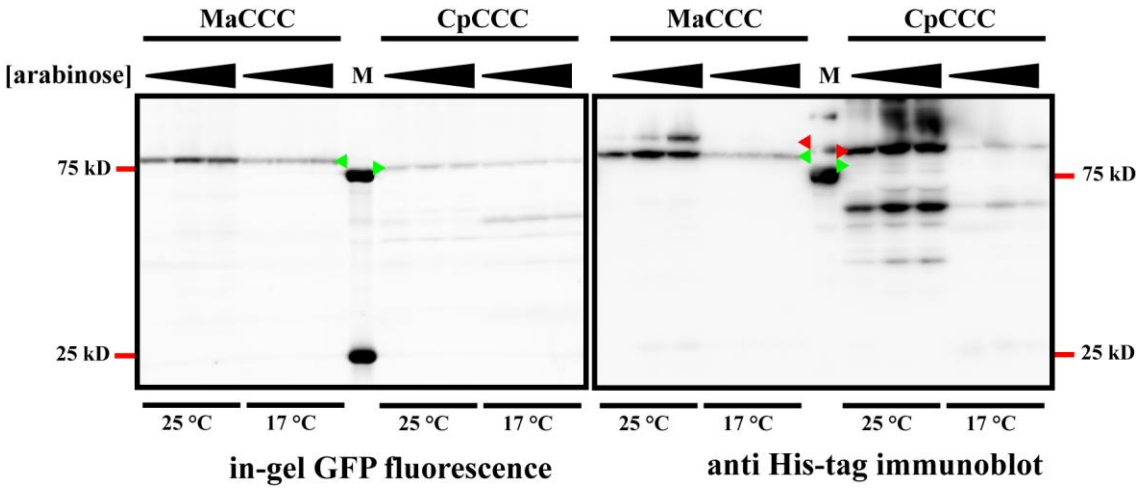


Fig 2-8 Fine Tuning of Expression in *E. coli*
MaCCC and CpCCC were expressed in *E. coli* with different temperatures and with varied arabinose concentrations of 10^{-1}, 10^{-2}, 10^{-3} % [w/v]. Whole cells equivalent to 20 µg of total protein were loaded after lysis and expression was analyzed with the GFP folding assay using GFP fluorescence (left panel) and anti His immunoblot (right panel). Green triangles correspond to well-folded, fluorescent protein whereas red triangles indicate non-folded protein. The weight is labeled corresponding to the marker (M). Both analyses were conducted using the same SDS-PAGE gel.

The constructs with high expression levels were selected for a fine-tuning of expression by varying the arabinose concentration and temperature. The protein quality and quantitiy was again monitored by GFP fluorescence and anti His-tag immunoblotting. MaCCC and CpCCC are shown as examples because both proteins show a different dependence on the inducer concentration and temperature (Fig 2-8). MaCCC showed a strong arabinose concentration-dependent expression behavior. The in-gel fluorescence signal increases with increasing arabinose concentration at 25 °C. In contrast at 17 °C the amount of expressed MaCCC stayed constantly low. For CpCCC the expression monitored by in-gel fluorescence stayed low at both temperatures. The analysis of expression using anti His-tag immunoblot illustrates the effect of overexpression on protein quality. MaCCC shows increasing expression of folded protein (lower band, green triangle) with higher inducer concentrations but additionally at high inducer concentration (10^{-1} % [w/v] arabinose) half of the expressed protein is in a misfolded state (higher band, red triangle). A concentration of 10^{-2} % [w/v] arabinose was thus the inducer concentration that was used for the scale-up of expressions cultures. CpCCC was mostly expressed as misfolded protein. While in-gel fluorescence still showed minor expression of folded protein at 25 °C the anti His-tag immunoblot revealed the highly aggregated state of the majority of expressed protein.

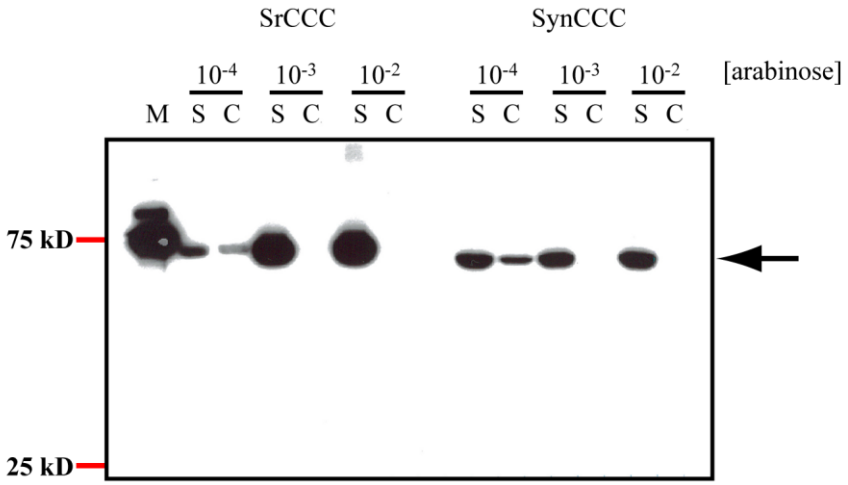


Fig 2-10 Expression of CCCs in *E. coli* with C-terminal His-tag Followed by Detergent Sedimentation Assay
SrCCC and SynCCC were expressed in *E. coli* with a C-terminal 3C site and a His_{10}-tag at different arabinose concentrations. Analysis of expression was performed by extraction of membrane proteins after cell lysis in 2 % DDM followed by a high-spin centrifugation. Solubilized cells (S) and supernatant after centrifugation (C) were analyzed by anti His-tag Western blot. The protein band is indicated by an arrow.

The overexpression CCC constructs with a C-terminal His-tag that lack GFP showed strikingly different results. The overall expression levels of the selected homologues were in general significantly lower and the sedimentation assay of detergent solubilized cells showed predominant expression of aggregated protein. Fine-tuning of the expression conditions by lowering the inducer concentration (e.g. 10^{-2}, 10^{-3} and 10^{-4} %

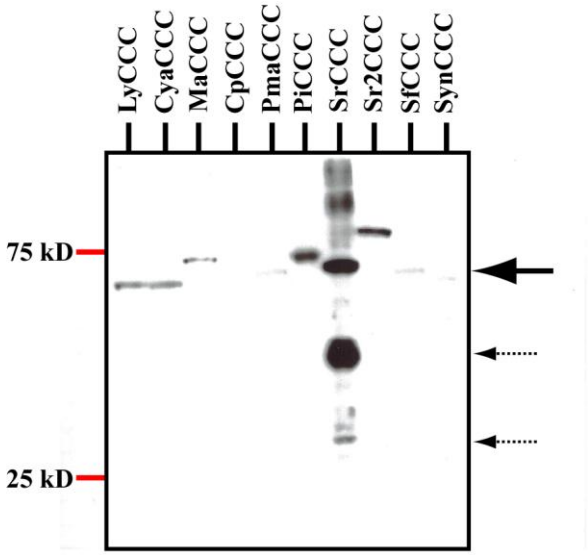


Fig 2-9 Expression Test in *E. coli* with N-terminal His-tag
Expressing homologues (arrow) harboring an N-terminal His tag were analyzed using anti His immunoblot. Equal amount of cells corresponding to 20 µg total protein were separated by SDS-PAGE. SrCCC showed strong degradation to lower molecular weight (dashed arrows).

[w/v] arabinose) resulted in the expression of low quantities of well-folded protein as judged from the fraction of the protein in the supernatant after high spin centrifugation (Fig 2-10). Among the GFP-less constructs only PiCCC and SynCCC showed reasonable expression of well-folded protein.

Besides proteins constructs carrying a C-terminal His_{10}-tag also constructs with a His_{10}-tag fused to the N-terminus of the protein were investigated. MC1061 *E. coli* cells containing CCCs with N-terminal His_{10}-tag were grown at 37 °C, induced by addition of 10^{-3} % [w/v] arabinose and incubated overnight at 25 °C. Equal amounts of cells were loaded on SDS-PAGE and analyzed by anti His-tag Western blot. The in-gel fluorescence analysis is not applicable in this case since these constructs lack a C-terminal GFP fusion as folding indicator. SrCCC exhibited strong overexpression while two transporters, PiCCC, and Sr2CCC were moderately expressed (Fig 2-9). Among the proteins that showed high expression, SrCCC was heavily degraded with a major band at lower molecular weight. All other proteins were very little or not overexpressed at all

Homologue	**Expression strain**	**Expression level P_{BAD}**		
		N-His	**C-His**	**C-His GFP**
PmaCCC	MC1061	-	+	+++
SfCCC	MC1061	-	-	-
SrCCC	MC1061	+++ *	+	-
CyaCCC	MC1061	+	- **	-
CwCCC	MC1061	-	-	+
PiCCC	MC1061	++	++	+++
TeCCC	MC1061	-	-	-
LyCCC	MC1061	+	-	+
SynCCC	MC1061	-	++	++
CpCCC	MC1061	-	-	- **
MaCCC	MC1061	+	++	+++
Sr2CCC	MC1061	++	- **	+

Tab 2-2 Summary of Expression Tests in *E. coli* Using the Arabinose Induction System
The prokaryotic homologues tested for expression are listed with their expression behavior in *E. coli*. N-terminal His-tag fusion (N-His), C-terminal His-tag fusion (C-His) and C-terminal GFP and His-tag fusion (C-His GFP) were analyzed. The expression was classified to good (+++), medium (++), low (+) and no (-) expression. Some homologues showed degradation (*) and others could not be solubilized during detergent sedimentation assay (**).

and thus not further analyzed.

2.3.2.2 Expression in *E. coli* with T7 promoter

Beside the expression efforts of prokaryotic CCCs under the control of the P_{BAD} promoter I have also investigated the expression under the control of the conventional P_{LAC} promoter. By using the T7 expression system present in pET vectors from Novagen, the gene of interest was inserted behind the very strong promoter from *E coli* bacteriophage T7. In the absence of T7 polymerase the promoter is turned off. The strains used for expression contain a single copy of T7 polymerase in their genomic DNA (DE3) under the control of a Lac-UV5 *lac* promoter. This promoter is turned off as long as the *lac* repressor (LacI) is bound to the operator. After dissociation of LacI through binding of lactose or IPTG from the operator the transcription and translation of T7 polymerase starts and the expression of the gene of interest begins. Since the *lac* promoter is leaky, basal expression level of T7 polymerase in the absence of inducer always leads to low expression of the protein of interest. Additionally, the expression is always strong in the presence of inducer without the possibility of regulation by a variation of inducer concentration. The main parameters screened during expression tests were thus temperature, media composition, expression time and the choice of the *E. coli* expression strains. The temperature was varied to investigate the effect of the cell metabolism on the expression behavior, since lower temperatures slow down the growth of the cells and thus decrease protein expression. The protein translation machinery thus has more time for correct membrane insertion and folding. Expression was screened at temperatures of 20 °C, 30 °C or 37 °C. The investigated media include standard LB medium, rich TB medium, minimal M9 medium and autoinduction medium (for details see materials and methods). The first three media differ in their richness, meaning amount and type of the carbon source and nitrogen content. TB medium and M9 medium are additionally buffered to maintain the pH during expression. The cells in all three media were induced by IPTG. In autoinduction medium the induction is triggered by a glucose/lactose mixture that regulates expression metabolically. As long as glucose is still present in the medium it supports the repression of the *lac* promoter since it decreases the cAMP level, which prevents the binding of the transcriptional activator catabolic activator protein (CAP). After glucose is metabolized by the growing cells, lactose as primary carbon source is imported which in turn slowly induces expression by binding to the repressor protein LacI.

Several investigated homologues showed detectable yields when expressed under the control of the T7 promoter. For two of the expressed transporters, the homologues SynCCC and MaCCC, the use of autoinduction medium significantly increased the expression compared to standard LB medium (Fig 2-11). This increase multiplied the

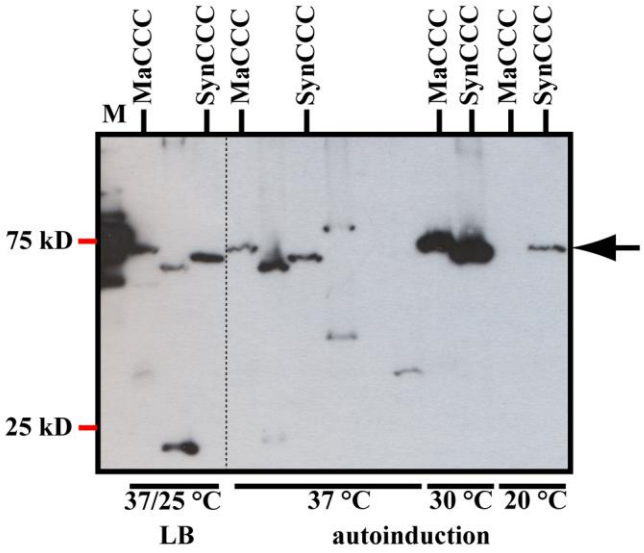


Fig 2-11 Expression Test in *E. coli* Using Different Media
MaCCC and SynCCC were tested for expression in different media using the *E.* coli strain BL21 (DE3). After cell lysis using 5 M urea 20 µg total protein were loaded per lane and detected with anti His Western blot. Cultures growing in LB at 37 °C to OD_{600nm}=0.5 and induction with 0.2 mM IPTG at 20 °C overnight were compared to culture growing at different temperatures overnight (37 °C, 30 °C, 20 °C) using autoinduction media.

yield of expressed protein by at least one order of magnitude. The temperature for optimal expression was found to be best at 30 °C.

After this initial test a time course was performed to analyse the progress of expression for MaCCC and SynCCC. This time series revealed an expression of 18 to 20 h as optimal. At 37 °C the highest relative expression was also reached after 18-20 h but the yield was lower than at 30 °C. 20 °C was included in the tests to investigate, whether longer expression increases the amount of protein. After 26 h at 20 °C weak expression was detected that was several orders below the yield of expression at 30 °C. A temperature of 30 °C with a growing time of 18-20 h was thus chosen for scale-up expression for subsequent purification (see next chapter).

Beside MaCCC and SynCCC, another homologue, SfCCC, showed high expression in TB medium at 37 °C when induced with 0.2 mM IPTG for 3 h at 37 °C. This homologue expresses about ten times more than MaCCC expressed in autoinduction medium. Several other homologues also showed decent expression levels at standard conditions illustrated in table Tab 2-3 .From the expression tests no information was obtained whether the protein is folded and indeed several of the homologues turned out to be not stable in detergent solution after purification, which will be described in a later chapter.

Homologue	Expression strain	condition	Expression level T7
SfCCC	BL21	TB, 37/25 °C o/n	+++
SrCCC	BL21	TB, 37/25 °C o/n	+++
CyaCCC	BL21	TB, 37/25 °C o/n	+++
CwCCC	BL21	TB, 37/25 °C o/n	+++
PiCCC	BL21	TB, 37/25 °C o/n	+++
TeCCC	BL21	TB, 37/25 °C o/n	+++
LyCCC	BL21	TB, 37/25 °C o/n	+++
SynCCC	BL21	Auto, 30 °, o/nC	++
CpCCC	BL21	TB, 37/25 °C o/n	-
MaCCC	BL21	Auto, 30 °C	++

Tab 2-3 Overview of Expression Using T7 Promoter in *E. coli*
Homologues were expressed in *E. coli* using BL21 (DE3) cells at the indicated condition. The expression level was classified as good (+++), medium (++), low (+) and none (-). 37/25°C o/n = growing at 37 °C to an OD_{600nm}=0.5 and induction with 0.2 mM IPTG at 25 °C overnight. Auto = autoinduction media

2.3.2.3 Expression in *L. lactis*

Analogous to the expression under the control of the P_{BAD} promoter the heterologous overexpression of all prokaryotic CCCs was tested in the Gram-positive bacterium *L. lactis* under the control of the P_{NIS} promoter. The parameters for tuning *L. lactis* expression are not as broad as for *E. coli*, besides the variation of the inducer nisin, a peptide of 34 amino acids, that is produced by *L. lactis* itself. *L. lactis* is sensitive to the growth medium (it grows best in M17 medium) and has a narrow temperature window. The temperature ranges between 20 °C with very high doubling time to a maximum of 30 °C. Initial expression tests were carried out at 20 °C overnight and induction with 10^{-3} % [v/v] of nisin.

Like in the case of the P_{BAD} promoter, the expression of C-terminal GFP fusion products was monitored by recording GFP fluorescence. The in-gel GFP fluorescence showed strong signals for CyaCCC, PmaCCC and PiCCC. Lower fluorescence was recovered for CwCCC, MaCCC and SrCCC. In no cases degradation was observed during expression. In-gel fluorescence of whole cell samples showed additionally to the protein of interest three bands from endogenous *L. lactis* proteins (Fig 2-12, dashed arrow).

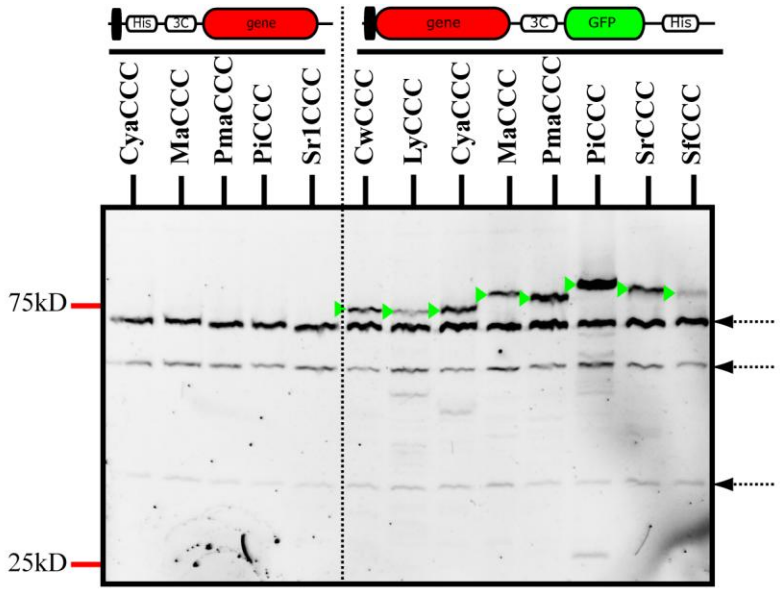


Fig 2-12 Analysis of Expression in *L. lactis* with In-Gel GFP Fluorescence
Selected homologues expressing in *L. lactis* were comparatively analyzed with in gel GFP fluorescence. Whole cell extracts were analyzed after cell lysis. 10 µg of total membrane protein was loaded on each lane. Both N- and C-terminal His-tag constructs were included. The green triangle indicates the position of the GFP fluorescence signal. The N-terminal fusions (first 5 lanes) lacking GFP served as control for *L. lactis* proteins showing fluorescence (dashed arrows).

The anti His-tag immunostaining showed additional bands of slower migrating protein, which, in contrast to *E. coli* , in *L. lactis* are not indicative for misfolded protein (Fig 2-13). Besides the expression of C-terminal GFP fusion proteins, all CCCs were also investigated with a N-terminal His-tag. Selected samples of transporters are shown for comparison in Fig 2-13. SrCCC showed strong expression as N-terminal His-tag construct. Other homologues like CwCCC, MaCCC, PmCCC and PiCCC did either not express or expressed very little.

Homologue	Expression strain	condition	Expression level pNZ L. lactis	
			N His	C His
CyaCCC		30 °C	-	+
MaCCC		30 °C	-	+
PmaCCC		30 °C	-	++
PiCCC		30 °C	-	+++
SrCCC		30 °C	+++	++
CwCCC		30 °C	-	+
LyCCC		30 °C	-	-
SfCCC		30 °C	-	-

Tab 2-4 Overview of the Expression Tests in *L. Lactis*
The expression level in *L. lactis* using a N-terminal His-tag or a C-terminal His-tag and GFP fusion classified as high (+++), medium (++), low (+) and minor or absent (-). These classifications are relative to the expression of the other homologues.

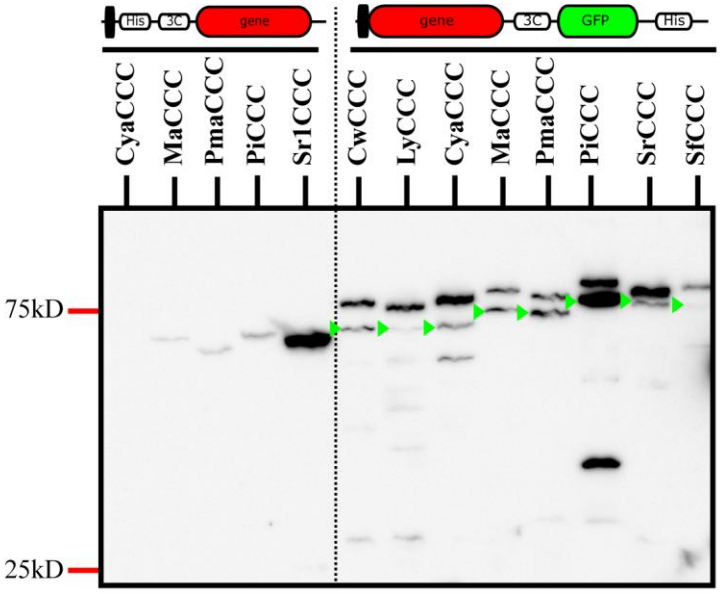


Fig 2-13 Analysis of Expression in *L. lactis* with Immunostaining
Selected homologues expressing in *L. lactis* were comparatively analyzed with anti His-tag immunoblot. Whole-cell extracts containing 10 µg total membrane protein were loaded after cell lysis. Both N- and C-terminal His-tag constructs were analyzed. The green triangle indicates the position of the GFP fluorescence signal (seeFig 2-12).

2.3.2.4 Conclusions of the Expression Tests

After extensive screening for expression conditions for twelve prokaryotic CCCs, some proteins could be identified as potential targets for scale-up and purification. Most of the homologues showed expression in *E. coli* under the control of the *lac*-T7 promoter system. This could have various reasons: The T7 promoter is a very strong promoter, that is continuously transcribed and might in that way foster the overproduction of proteins. The initial tests with the T7 system didn`t allow to conclude on the quality of the expressed material. The pBAD plasmids used with a C-terminal GFP fusion gave the opportunity to also assay the quality of the produced material. Most of the expressed proteins in this system were well-folded or contained a major fraction of folded material.

The use of *L. lactis* as different expression host showed in general lower expression but allowed the identification of 2 additional proteins, CyaCCC and SrCCC whose investigation was pursued.

Homologue	**T7**	**pBAD-N**	**pBAD-C**	**pNZ**
PmaCCC	-	-	+/+++	-/++
SfCCC	+++**		-/-	-/-
SrCCC	+++**	+++*	+/-	+++/++
CyaCCC	+++**	+	-**/-	-/+
CwCCC	+++**	-	-/+	-/+
PiCCC	+++**	++	++/+++	-/+++
TeCCC	+++**	-	-/-	-/-
LyCCC	+++**	+	-/+	-/-
SynCCC	++	-	++/++	-/-
CpCCC	-	-	-/-**	-/-
MaCCC	++	+	++/+++	-/+
Sr2CCC	+++**	++	-**/+	-/-

Tab 2-5 Summary of the Expression Test in *E. coli* and *L. Lactis*
The prokaryotic homologues tested for expression are listed with their expression behavior in *E. coli*. Using the T7 promoter only C-terminal His-tags are listed, for P_{BAD} the N-terminal tag (pBAD-N) and the C-terminal tag (pBAD-C) including only His-tag or with GFP (x/x). The expression in L. lactis I indicated for N- and C-terminal tag (x/x). The expression was classified to good (+++), medium (++), low (+) and no (-) expression relative to each other in the used expression system. Some homologues showed degradation (*) and others could not be solubilized during detergent sedimentation assay (**).

The broad screening of expression condition reduced the choice of possible proteins for scale-up and purification from twelve to five. In the next step, the biochemical characterizations of the transporters will decreases the pool of suitable candidates further. The following chapter describes the biochemical characterization of the preselected CCCs with respect to their extraction behavior and stability in different detergents. The major quality check during extraction and purification was SEC elution profile. The GFP fluorescence was used for detection of appropriate constructs. The different behavior during extraction and purification will be discussed in detail for the selected homologues MaCCC, SfCCC and SynCCC. MaCCC was the homologue with the most promising behavior and was thus most thoroughly characterized.

2.3.3 Purification of Prokaryotic CCCs

The purification of membrane proteins starts with the extraction of the protein from the membrane with various detergents (Fig 2-2). After successful solubilization of the His-tagged protein the purification proceeds by immobilized metal affinity chromatography (IMAC), and optional removal of the His-tag by limited proteolysis at an introduced protease cleavage site and size exclusion chromatography (SEC).

IMAC takes advance of the chelating properties of Me2+ ions. The C-terminally fused His_6-tag or His_{10}-tag can bind to Ni^{2+}, Co^{2+}, Fe^{2+} and Zn^{2+} ions with decreasing affinity. Thus protein can be retained on Me^{2+} chelating matrix. Different resins with varying properties were used during this study: Ni-NTA (Quiagen), Co-Talon (Clontech), Ni-, Co-, Fe-, Zn-sepharose (Amersham) and Ni-IDA (Macherey-Nagel). Additionally the binding properties of the protein vary depending on the buffer composition, tag accessibility and tag length. Imidazole, which mimics the side chain of histidine, is a competing agent for protein binding and is used for the removal of non-specifically bound proteins at lower concentrations and the elution of the protein from the resin at higher concentration. To determine the boundaries for optimal washing and elution of the protein a step gradient of different imidazole concentrations was used.

For continuing analysis, the His tag was removed from the protein by limited proteolysis using the 3C precision protease that cleaves a 10 residue long recognition sequence that was introduced between the protein and the tag (Fig 2-14). The cleavage removes the flexible His-tag, which is usually beneficial for crystallization. Moreover,

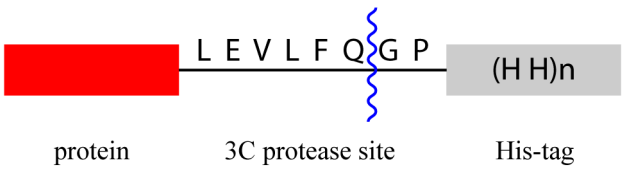


Fig 2-14 3C Protease Site
The protein of interest (red) is followed by a protease site and the His-tag (grey). After digestion with 3C protease (blue) six residues remain on the protein as short C-terminal tail.

since the His-tag is partially charged at neutral buffer conditions it might interfere with the formation of crystallization contacts. After cleaving of the tag a tail of six residues remain attached to the C-terminus of the protein.

The most important tool in this study to examine the mono-dispersity and oligomeric behavior of the purified transporter was size exclusion chromatography. From this technique the distribution of oligomeric states of the protein and the amount of aggregation can be derived in a simple straightforward manner. Additionally SEC offers

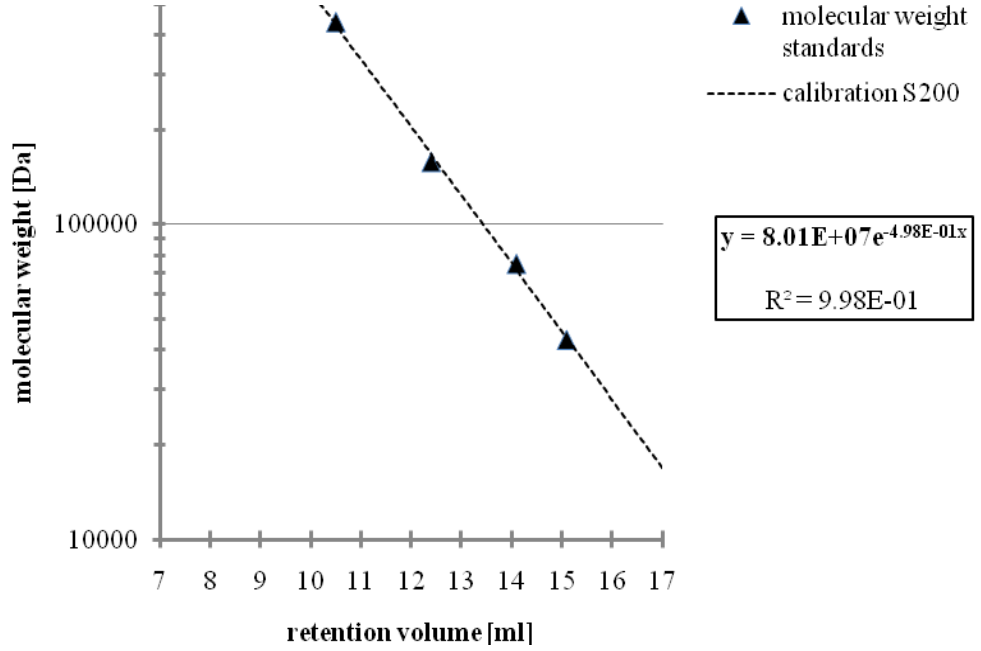


Fig 2-15 Calibration Curve of S200 GE Healthcare Column
The S200 column used in this study was calibrated using standard molecular weight markers. The retention volume was measured on an Äkta Prime purification system that was used for all purifications.

the possibility to separate impurities with very different molecular weight from the target protein.

SEC separates the proteins based on their size and shape. A calibration curve was recorded with known protein standards to obtain the relationship between molecular weight and the retention volume of the column (Fig 2-15). With the assumption of globular proteins, the molecular weight for the tested protein can be estimated. Since membrane proteins are surrounded by a detergent micelle (protein detergent complex, PDC) in solution their apparent molecular weight is higher than expected for protein alone because of the contribution of the detergent.

2.3.3.1 SfCCC

For scale-up SfCCC was expressed under different conditions to maximize the expression level and the fraction of extracted protein. The protein was expressed under the control of the T7 promoter. Due to low expression in autoinduction media (which gave optimal expression for MaCCC) SfCCC was expressed in TB media (20 °C o/n, 28 °C o/n and 37 °C for 3 h). Additionally the inducer concentration varied between 0.25 to 1 mM IPTG. At strong induction conditions (1 mM IPTG, 37 °C) SfCCC was mostly found in the pellet after solubilization and high spin centrifugation. Only Fos-cholin 12 could extract a minor fraction from the membrane and keep it in solution (Fig 2-16A). With the shift to milder expression conditions (0.5 mM IPTG, 20 °C, o/n) the

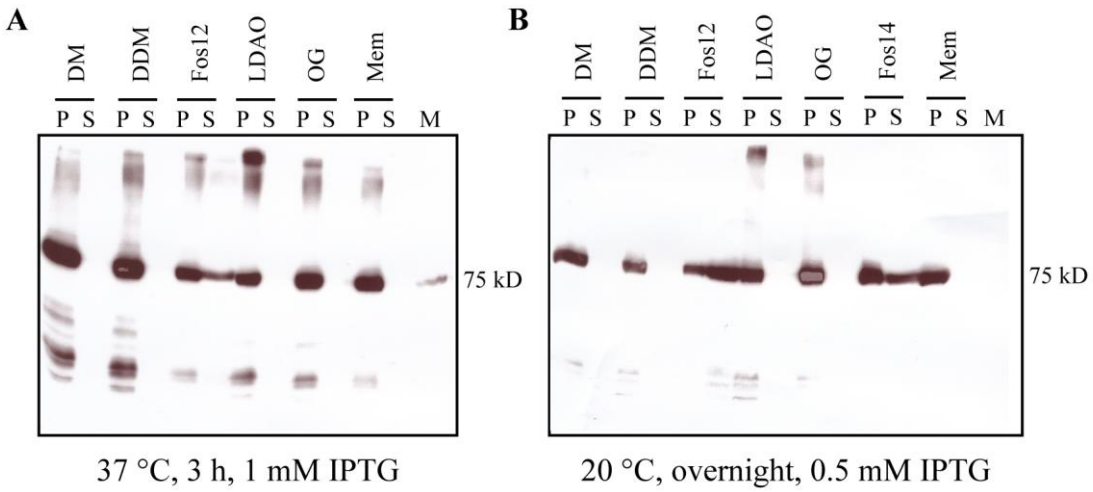


Fig 2-16 SfCCC Extraction Tests in Dependence of Expression Conditions
SfCCC was expressed at different temperatures and inducer concentrations. After preparation of the membranes from expression cultures at conditions listed below the panel, samples were incubated with 2 % of the indicated detergent followed by high-spin centrifugation. Pellet (P) and supernatant (S) of the centrifugation were analyzed by SDS-PAGE and anti His-tag Western blot to quantify the extraction efficiency. **A** SfCCC was expressed at 37 °C and induced with 1 mM IPTG for 3 h. Only with Fos12 a minor fraction could be extracted from the membranes. **B** With expressing SfCCC at 20 °C overnight and induction with 0.5 M IPTG most of the protein could be extracted with Fos12 and half with Fos14. DM, decyl maltoside; DDM, dodecyl maltoside; Fos12, Fos-cholin 12; LDAO, lauryl dimethylamine n-oxide; OG, octyl glucoside; Fos14, Fos-cholin 14

expression was improved. Fos-cholin 12 was able to extract most and Fos-cholin 14 about 30 % of the protein (Fig 2-16B). The choice of these two detergents in combination with the milder expression conditions were thus used to proceed with purification.

Since Fos-cholins are very potent detergents the initial solubilization concentration of 2 % (w/v) was decreased to 0.25-0.5% (w/v). The following IMAC purification on Ni-NTA agarose in Fos-cholin 12 using the His_{10}-tag was initiated with batch binding in the presence of 15mM imidazole. The resin was subsequently washed on column with 50mM imidazole and the bound protein was eluted with 300mM imidazole. SfCCC showed a small amount of leakage during washing and could be recovered in a very

pure state in the elution fraction (Fig 2-17D). The His_{10}-tag was successfully removed by incubation with 3C protease, and the following SEC using a S200 column showed a peak at 11.3 ml elution volume, which corresponds to a molecular weigth of 290 kD that would hint at an oligomeric state higher than a monomer (Fig 2-17A). A significant amount of the protein was eluting in the void volume of the SEC column (arrow) indicating the presence of higher molecular aggregates. This presence of higher oligomers could indicate instability of the protein or could also be due to Fos-cholin 12 extracting previously misfolded protein from the membrane. The stability of SfCCC over time (1 and 5 days) was investigated by reinjection of the main SEC elution peak at later time points. Since the reinjected protein did not show aggregation, it can be concluded that it is reasonable stable in detergent solution (Fig 2-17B and C). The elution volume stayed constant over time indicating that the oligomeric assembly did not change over time.

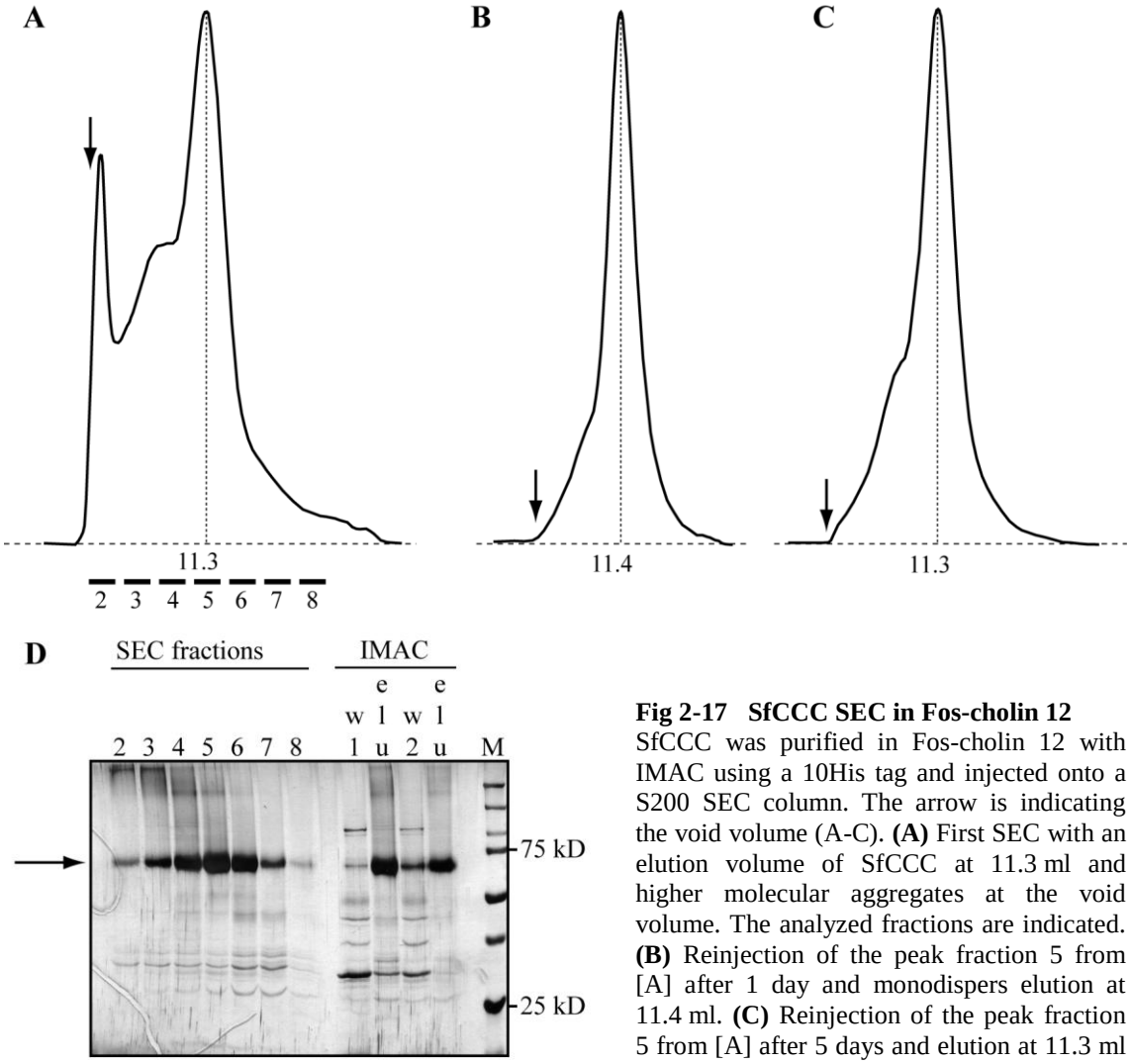


Fig 2-17 SfCCC SEC in Fos-cholin 12
SfCCC was purified in Fos-cholin 12 with IMAC using a 10His tag and injected onto a S200 SEC column. The arrow is indicating the void volume (A-C). **(A)** First SEC with an elution volume of SfCCC at 11.3 ml and higher molecular aggregates at the void volume. The analyzed fractions are indicated. **(B)** Reinjection of the peak fraction 5 from [A] after 1 day and monodispers elution at 11.4 ml. **(C)** Reinjection of the peak fraction 5 from [A] after 5 days and elution at 11.3 ml with a slight shoulder. **(D)** SDS-PAGE of the IMAC and SEC fraction from [A] and silver staining. SfCCC band is indicated by an arrow. The protein shows a clean elution from the IMAC column [elu] and clean SEC appearance. The IMAC 50 mM

imidazole wash caused minor leakage of SfCCC [w1,w2]. The retention volume indicated in [ml] is plotted against the absorption at 280 nm

The stability of SfCCC in different detergents was investigated by dialysis of the main protein peak after SEC of protein purified in Fos-cholin 12. Aliquots of the protein were dialyzed overnight at 4 °C into the detergents DM, DDM, LDAO, $C_{12}E_9$ at a concentration exceeding 2 times the respective CMC. After dialysis the samples were subjected to SEC with the column equilibrated with running buffer containing solely the new detergent. The detergent exchange experiments revealed that SfCCC is instable in most investigated detergents. Only DM and LDAO allowed the recovering of a significant part of the protein at a symmetric elution peak (Fig 2-18). In all tested detergents the formation of high molecular aggregates was observed. In DDM the aggregates overlap with the main elution peak. $C_{12}E_9$ was not tested by gel-filtration because the protein precipitated during overnight dialysis. The retention volume of the main peak shifted during the detergent exchange by up to 1 ml from 11.3 ml to 12.2 ml (DM). This shift in the molecular weight (which corresponds to n apparent change in the molecular weiht of 185 kD) could either be due to a change in the micelle size or it could indicate a change in the oligomeric state of the protein.

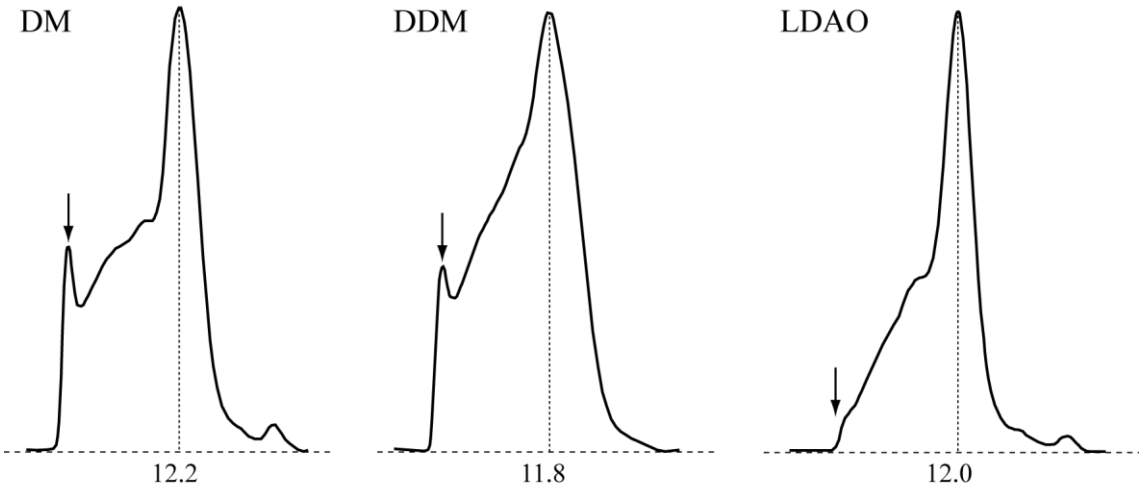


Fig 2-18 SfCCC Detergent Stability Screen
After purification of SfCCC with the detergent Fos-cholin 12 including IMAC and SEC the main peak fraction was dialyzed overnight in different detergents. A SEC with the new detergent using a S200 column followed. In DM and DDM the protein showed moderate stability with high molecular aggregates also in the void volume (arrow). In LDAO the protein was more stable without heavy aggregation. Overall the elution volume increased up to one milliliter compared to the Fos-cholin 12 elution hinting for a oligomeric rearrangement to a monomer. $C_{12}E_9$ showed heavy precipitation and was not tested on SEC. The retention volume indicated in [ml] is plotted against the absorption at 280 nm.

In summary, SfCCC could be expressed using the T7 promoter with a high yield, while the expression under the control of the P_{BAD} promoter was unsuccessful. Biochemical investigations identified detergents of the Fos-cholin family to be the only detergents suitable for extraction of SfCCC from the membranes. After separating high

molecular aggregates with SEC the main elution peak, corresponding to a putative dimeric transporter, was found to be stable in solution. Despite an overall promising behavior crystallization trials of SfCCC in the presence of Fos-cholin 12 and 14 , DM, DDM and LDAO and with the addition o lipids didn't show crystalline behavior.

2.3.3.2 SynCCC

SynCCC was one of the first proteins that could be identified using a *blastp* search at the beginning of the thesis in 2002. The first expression trials were carried out in parallel with MaCCC and the expression in *E. coli* growing in autoinduction media overnight at 30 °C was identified to yield the highest amounts of protein. The extraction from isolated membranes showed Fos-cholin 12 and 14 to be the only detergents that efficiently solubilize SynCCC.

The first purifications of SynCCC were conducted using the C-terminal His_6-tag for IMAC. After batch binding to Ni-NTA agarose resin and washing in the presence of 15 mM imidazole the protein was eluted with 300 mM imidazole. The following SEC showed a characteristic elution profile that indicates the presence of different defined oligomers. After a small fraction of aggregated protein (arrow) eluting after the void volume the main elution consisted of a double-peak with elution volumes of 10.8 ml and 11.5 ml corresponding to an apparent molecular weight of 370 kD and 260 kD,

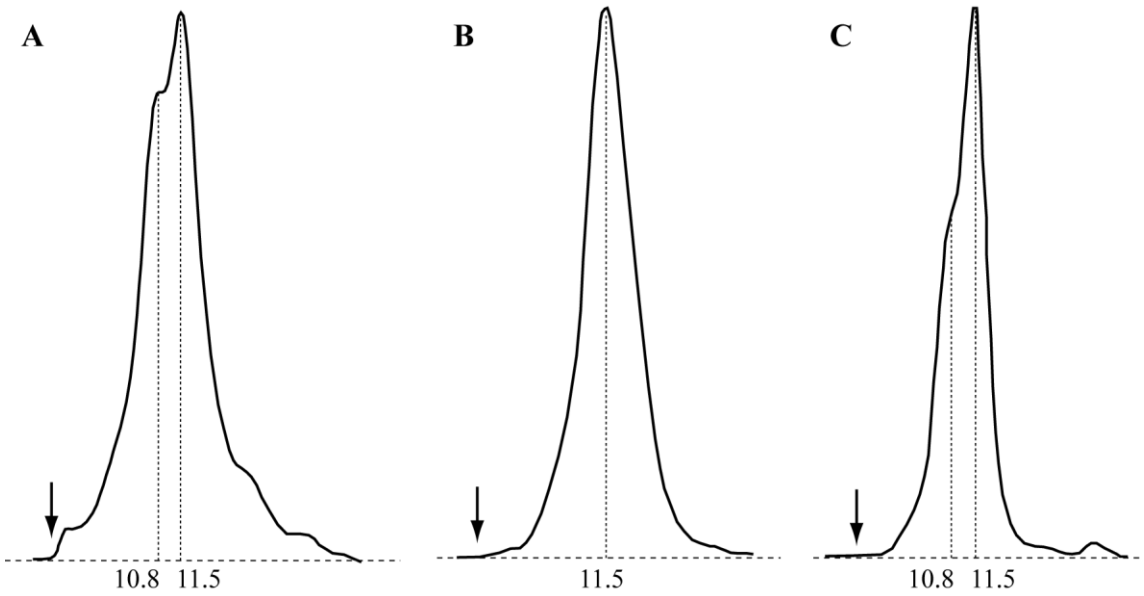


Fig 2-19 SynCCC SEC Behavior and Stability Test in Fos-cholin 12
SynCCC was purified using Fos-cholin 12 and subjected to SEC on a S200 column. The retention volumes are indicated in [ml] and the void volume by an arrow. **(A)** First injection after IMAC purification and cleavage of the His-tag. SynCCC eluted as double peak with elution volumes of 10.8 and 11.5 ml. A reinjection of both elution peaks 24 h later resulted in the predominant elution at higher volume ([A]-10.8 ml in **(B)**; [A]-11.5 ml in **(C)**). The retention volume indicated in [ml] is plotted against the absorption at 280 nm

respectively (Fig 2-19A). After reinjection of fractions of the two peaks a single monodisperse peak at a higher volume was recovered (Fig 2-19B and C).

A broad crystallization screen with both peaks in Fos-cholin 12 did not show any hints of crystallization.

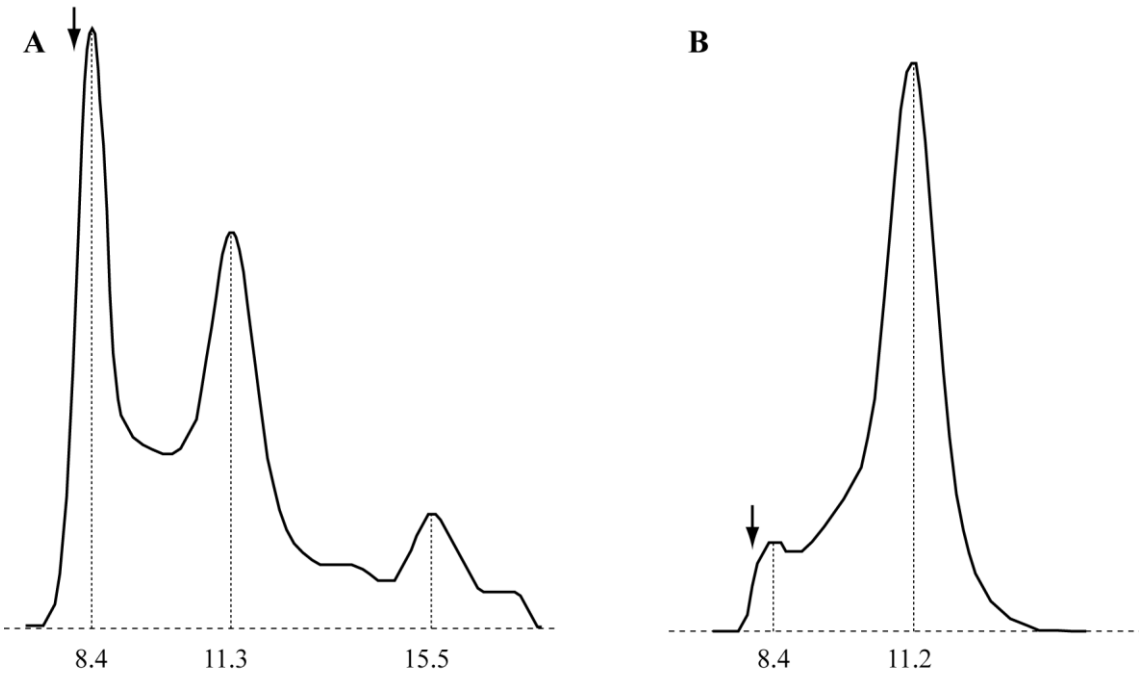


Fig 2-20 SynCCC SEC Behavior and Stability
SynCCC was purified in DDM with IMAC using the His_{10}-tag. Before injection onto a S200 column the C-terminal His-tag was removed. The retention volume is indicated in [ml]. **(A)** The SEC chromatogram showed an elution at 11.3 ml corresponding to molecular weight of a dimer. Additionally heavy aggregation could be observed eluting right after the void volume (arrow). **(B)** Reinjection of the main elution peak of [A] 24 h later. A stable dimer was recovered (11.2 ml) with little aggregation tendency. The retention volume indicated in [ml] is plotted against the absorption at 280 nm

In a second expression strategy using the P_{BAD} expression system combined with a C-terminal GFP fusion the expression of SynCCC was optimized for maximum expression of well-folded protein. Nevertheless the yield of purified protein after expressing overnight at 25 °C with an induction of 0.0001 % [w/v] arabinose was still lower than the expression obtained using the T7 promoter. Interestingly SynCCC after optimized P_{BAD} expression showed a different extraction behavior compared to the expression with the T7 promoter since the protein was now extractable in DDM. SynCCC containing a His_{10}-tag was purified in DDM by binding to Ni-NTA in batch in the presence of 15 mM imidazole. The protein was washed with 50 mM imidzole and eluted with 300 mM imidazole. The following SEC using a S200 column is partly aggregated but otherwise showed a single peak at 11.3 ml (Fig 2-20A). The elution volume corresponds to a molecular weight of 290 kD, which indicates a dimeric behavior. The aggregates could be successfully removed by SEC and a reinjection of the

main elution peak preserved the apparent dimeric behavior of the protein (Fig 2-20B). Although the protein expressed under the control of the P_{BAD} promoter showed an improved detergent stability, crystallization trials in DDM (with and without addition of lipids) were unsuccessful.

SynCCC could be expressed under the control of two different promoters. When expressed with the T7 promoter system in BL21 (DE3) cells, the protein could only be extracted and purified in Fos-cholin 12. The SEC showed two overlapping peaks indicating two stable oligomeric species. Expression with the P_{BAD} promoter in MC1061 cells in contrast generated a protein that was extractable in DDM. A symmetric, monodisperse peak could be recovered after reinjection. Both preparations did not show any sign of crystallization.

2.3.3.3 MaCCC

Similar to SynCCC MaCCC previously showed promising expression behavior in both *E. coli* promoter systems T7 and P_{BAD}. After expression under the control of the T7 promoter in autoinduction medium the solubilization of the protein from isolated membranes in different detergents was investigated. A broad set of detergents was chosen, which was widely used for purification and crystallization including the maltosides with different chain length, DM and DDM, LDAO, $C_{12}E_9$ and Fos-cholin 12.

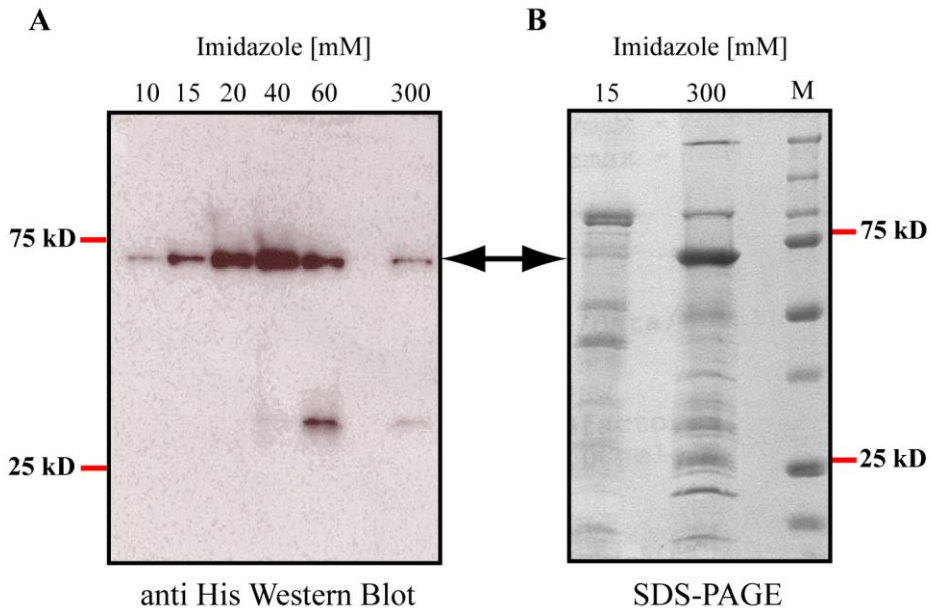


Fig 2-21 MaCCC-His_6 IMAC Purification Test
After batch binding of DM solubilized MaCCC (with a His_6-tag) to equilibrated Ni-NTA a stepwise imidazole gradient was performed to investigate the elution properties and thus determine the maximal washing concentration. **(A)** Anti His-tag Western blot analysis of the wash fraction with 10, 15, 20, 40, and 60 mM Imidazole. The elution of MaCCC (arrow) already started at 15 mM and is dominent at 40 mM imidazole. **(B)** SDS-PAGE analysis and coomassie stain of the 15 mM wash fraction and the 300 mM imidazole elution fraction. At 15 mM little MaCCC (arrow) leaked and at 300 mM. MaCCC is the major protein band.

The supernatants and the pellets after high-speed centrifugation were analyzed with SDS-PAGE followed by anti His-tag immunoblotting that showed DM and Fos-cholin 12 to have the best extraction properties.

DM solubilized MaCCC was used for the subsequent purification steps. For MaCCC-His_6, Ni-NTA showed the best binding properties. A step gradient from 0 mM to 60 mM imidazole controlled by anti His-tag Western blot indicated 15 mM imidazole to be the limiting concentration for washing the resin without losing significant amount of target protein (Fig 2-21A, arrow). The protein was eluted with 300 mM imidazole and analyzed by SDS-PAGE. Even though many impurities could be removed during washing a significant amount of impurities is still present in the elution fraction (Fig 2-21B).

To improve the binding of MaCCC to resin and thus enhance the purity of the sample, MaCCC was fused to a C-terminal His_{10}-tag. A His_{10}-tag binds stronger to Ni-NTA and thus allows more stringent washing steps, which result in cleaner protein after elution. A step-gradient after batch binding of MaCCC-His_{10} to Ni-NTA of 10 mM to 60 mM imidazole allowed washing off most impurities. The protein eluted at 300 mM imidazole. During washing with 60 mM imidazole the protein did not eluted from the

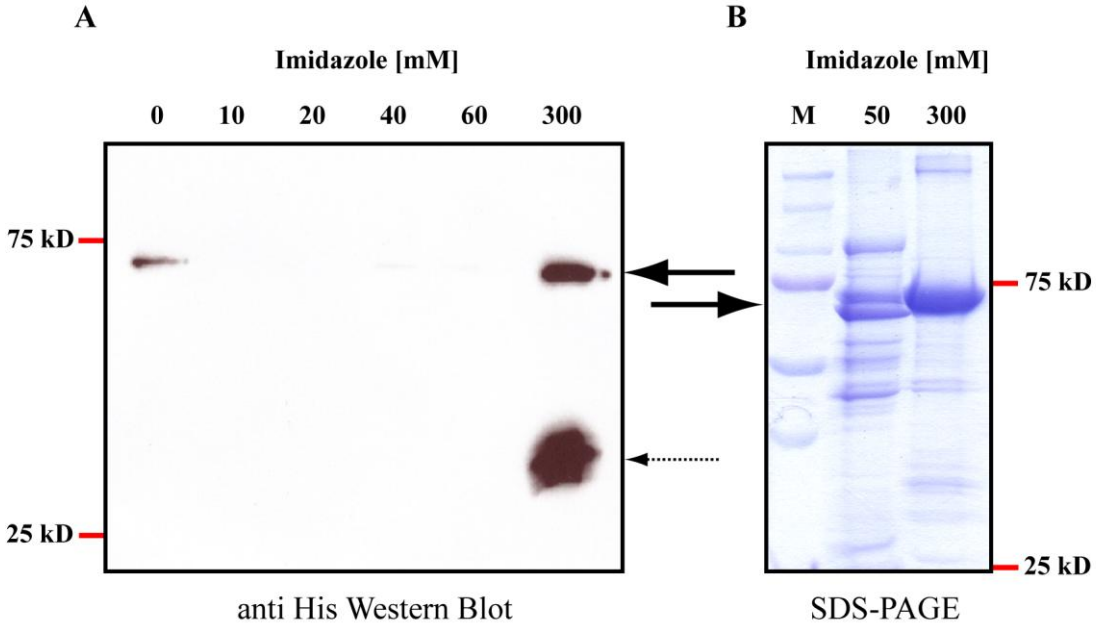


Fig 2-22 MaCCC-His_{10} IMAC Purification Test
After batch binding of DM solubilized MaCCC (with a His_{10}-tag) to equilibrated Ni-NTA a stepwise imidazole gradient was performed to investigate the elution properties and thus determine the maximal washing concentration. **(A)** Anti His-tag Western blot of the wash fractions with 0, 10, 20, 40 and 60 mM imidazole. At 0 mM only residual MaCCC from column loading was washed off. MaCCC eluted at 300 mM (arrow). The cleaved C-terminal domain still harboring the His-tag was co-purified (dashed arrow). **(B)** SDS-PAGE analysis and coomassie stain of the 50 mM wash fraction and the 300 mM imidazole elution fraction. At 50 mM most of the impurities were washed out to obtain a very clean 300 mM imidazole elution fraction of MaCCC (arrow).

Ni-NTA resin and could thus be quantitatively recovered with the high-imidazole elution (Fig 2-22A arrow). The domain of MaCCC was cleaved from the full-length protein and a fraction of the molecules was co-purified since it still contained a His-tag on its C-terminus (dashed arrow). The use of a His_{10}-tag for purification in binding to

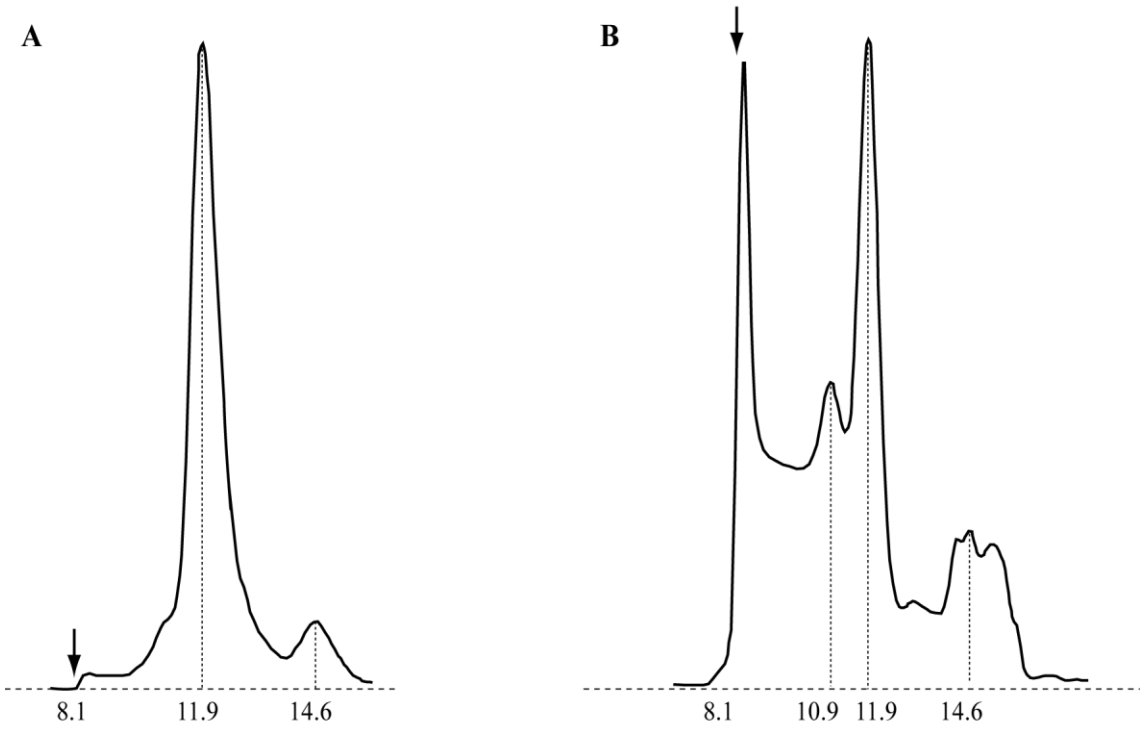


Fig 2-23 MaCCC SEC Profile
Purification of MaCCC in DM with a S200 SEC column. The void volume is indicated by an arrow. **(A)** At 4 °C MaCCC eluted at 11.9 ml which corresponds to an apparent molecular mass of 200 kD with minor aggregation at the void volume (arrow). At a volume of 14.6 ml the soluble C-terminal domain eluted which was co-purified as degradation product of the full-length MaCCC. **(B)** At room temperature a large aggregation peak at the void volume was visible (arrow). With the elution volumes of 10.9 ml and 11.9 ml an apparent molecular weight of 350 kD and 200 kD, respectively was observed. At a volume of 14.6 ml the soluble C-terminal domain eluted which was co-purified as degradation product of the full length MaCCC. The retention volume indicated in [ml] is plotted against the absorption at 280 nm

Ni-NTA affinity resin thus showed a major improvement in purity compared to MaCCC-His_6 (Fig 2-22B).

After IMAC purification the eluted protein was incubated for 2h at 4 °C with 3C protease at a molar ratio of 1mol of protease per 5 mol of protein. After this step a complete removal of the His tag was confirmed by the absence of an anti His-tag signal in a Western blot of MaCCC. Size exclusion chromatography was used to gain information about the homogeneity of the protein sample in solution. MaCCC was subjected to SEC using a S200 sepharose column. Runs at either 4 °C or 20 °C in the presence of DM were carried out to evaluate the stability at these temperatures. At 4 °C the size exclusion chromatogram tracing the absorbance at 280 nm showed an elution

peak at a volume of 11.9 ml (Fig 2-23A). This retention volume corresponds to the calculated mass of around 200 kD. Minor absorption was detected at the void volume where high molecular aggregates elute.

The purification at 20 °C in DM revealed a different behavior of the protein (Fig 2-23B). The aggregation peak directly after the void volume was very high and indicated a large fraction of high molecular mass aggregates in the region between 8-10 ml. The elution volume of 11.9 ml indicated the same molecular mass as at 4 °C, 200 kD. When comparing the two SEC profiles at the different temperatures it was obvious that the stability of the protein was improved at 4 °C. The effective absences of aggregates and a single, symmetric SEC peak show a monodisperse and stable protein that did also not show signs of aggregation after several days at 4 °C. This protein has thus very promising properties for crystallization.

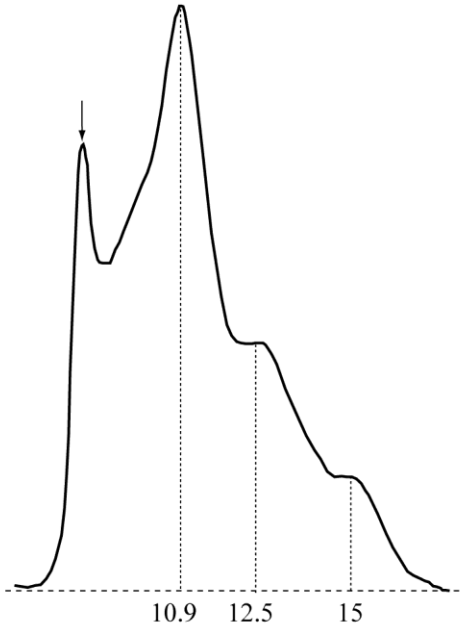


Fig 2-24 MaCCC Purification in Fos-cholin 12
MaCCC was solubilized and further purified in Fos-cholin 12. The SEC with a S200 column revealed a main elution at 10.9 ml corresponding to a dimer with a shoulder at 12.5 ml. Major protein quantities eluted directly after the void volume (arrow). The retention volume indicated in [ml] is plotted against the absorption at 280 nm

A second detergent tested for purification of MaCCC was Fos-cholin 12 since it also showed good membrane extraction properties. The preparation was carried out at 4 °C with the optimized parameters for IMAC mentioned above. The SEC elution profile in Fos-cholin 12 at 4 °C showed a different chromatogram than observed for DM (Fig 2-24). The main elution peak at 10.9 ml corresponds to an apparent molecular weight of 350 kD, 1.5 times higher than observed in DM. A shoulder as result of two overlapping peaks eluted at 12.5 ml (apparent molecular weight=160 kD). The high absorption at the void volume corresponds to protein aggregation. The reinjection of the main fraction around 10.9 ml recovered a single elution peak at the same volume. The observed

difference in the elution volumes of the different peaks of Fos-cholin 12 was probably due to different oligomeric states. The difference in the main peak retention volume between Fos-cholin 12 and DM either corresponded to different oligomeric states or a change in the elution properties due to the difference in the detergent micelle.

The detergent exchange from DM into other detergents was investigated to identify other protein detergent complexes for crystallization. After purification in DM aliquots of the main peak fraction of MaCCC after SEC were dialyzed overnight into different detergents. After 12-18 h dialysis the elution profiles of MaCCC were assayed in a column equilibrated in the detergent dialyzed against. In the detergents OM, NM, DDM and LDAO, MaCCC was reasonable stable and eluted at SEC at a volume of around 11.9 ml (213 kD) as single, symmetric peak (Fig 2-25). With decreasing chain-length of the hydrophobic tail OM<NM<DDM the chromatogram showed a partial cleavage of the C-terminal cytoplasmic (15.6 ml – 34 kD) from the full-length protein. In contrast to

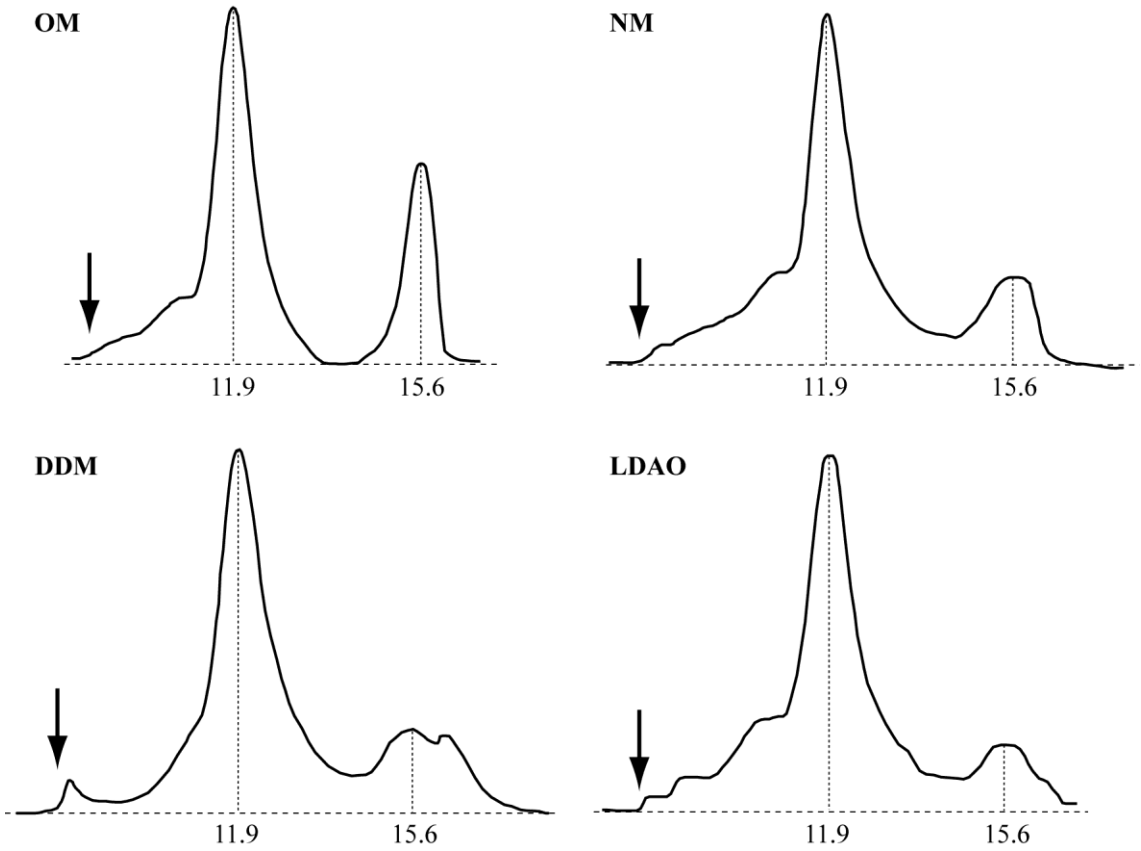


Fig 2-25 MaCCC Detergent Stability Test
SEC of MaCCC in different detergents. Aliquots of the main SEC peak of MaCCC in the presence of DM were dialyzed overnight into different detergents and subjected to SEC using a column equilibrated in the new detergent. The chromatograms show single peaks at a similar elution volume in OM, NM, DDM and LDAO. The void volume is indicated by an arrow. The retention volume indicated in [ml] is plotted against the absorption at 280 nm

maltosides and LDAO, the exchange into $C_{12}E_9$ and OG has led to the precipitation of the protein during dialysis.

Since the previous experiments indicated that MaCCC might dissociate into monomers I have investigated the stabilization of oligomeric protein by supplementation of lipids to the purification buffers. MaCCC-His_{10} was purified as previously described but with addition of *E. coli* polar lipids at a concentration of 0.1 mg/ml throughout the purification. After removal of the His_{10}-tag using the 3C protease the subsequent SEC with a S200 column showed a dominant elution peak at a

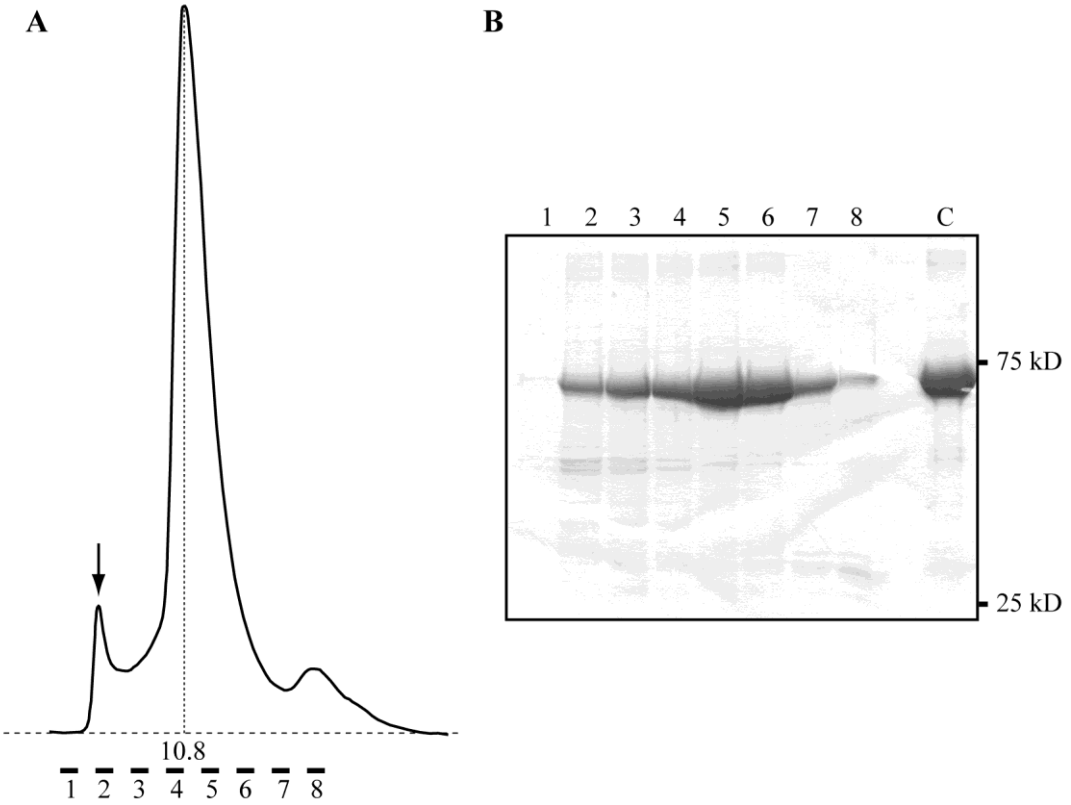


Fig 2-26 MaCCC-10His SEC in DM+Lipids
SEC of MaCCC in DM in the presence of lipids. **(A)** The chromatogram shows a symmetric elution peak at 10.8 ml with a little aggregate peak at the void volume (arrow). **(B)** The fractions 1-8 (bar in **A**) were analyzed by SDS-PAGE and coomassie stain. The pooled and concentrated fractions 5 and 6 are shown in lane C. The retention volume indicated in [ml] is plotted against the absorption at 280 nm

volume of 10.8 ml (Fig 2-26). This volume was shifted with respect of the elution volume in the absence of lipids by 1.1 ml and corresponds to an apparent molecular mass of 370 kD. The protein was stable for several days at 4 °C in the SEC buffer containing lipids and did not change its elution volume upon reinjection. With the improvement of the expression using autoinduction medium, the optimization of the IMAC protocol taking advantage of the His_{10}-tag and the addition of lipids during purification the overall yield was reproducibly around 200 µg purified protein/liter of expression culture.

2.3.3.3.1 The Oligomeric state of MaCCC

The observed difference in retention volumes in SEC in the absence and presence of lipids indicate a lipid-dependence of the oligomerization of the protein. This behavior was investigated in detail. The knowledge of the structural organization of the isolated C-terminal domain of MaCCC (see chapter 2.4.1) guided the mutation of residues in the proposed dimer-interface in full-length MaCCC. E505 and Q622 located in proximity in the proposed interface of the C-terminal domain were mutated to cysteines to allow a covalent crosslinking by disulfide bond formation. The oligomeric behavior was

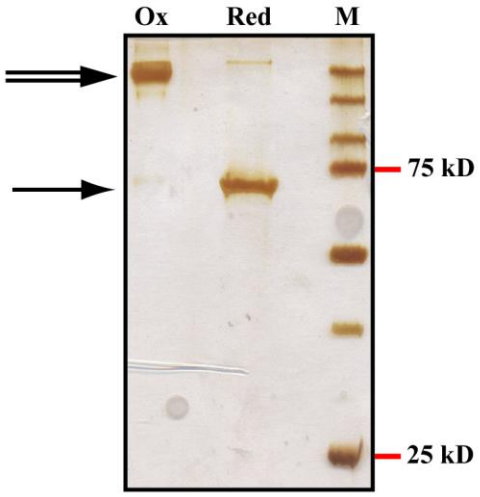


Fig 2-27 MaCCC Cysteine Crosslink
MaCCC E505Q622C was purified under oxidizing conditions and subjected to SDS-PAGE. Keeping the oxidative environment intact (Ox, left lane) led to a 250 kD migration behavior. Under reducing conditions in 0.5 mM TCEP decomposition of the cysteine bridges reconstitute the migration behaviour of the wild-type around 70 kD.

followed by SDS-PAGE and multi angle light scattering (MALS). After purification of MaCCC E505Q622C under oxidizing conditions (air O_2) the protein was applied to SDS-PAGE separation under oxidizing conditions (Ox) or to reducing conditions by addition of 0.5 mM TCEP (Red) (Fig 2-27). The oxidized mutant migrated around 250 kD, the molecular weight of a dimer. After reduction of the disulfide bridge the mutant showed a similar monomeric migration behavior as wild-type MaCCC (around 70 kD), which strongly suggests that MaCCC forms dimers. The result, however is only qualitative and should be taken with caution, since even a dimeric protein in equilibrium with its monomeric would quantitatively crosslink upon prolonged incubation.

The oligomeric state of MaCCC was also analyzed using SEC-MALS with UV and dRI detection (Slotboom, Duurkens et al. 2008). This method uses the scattering intensities of particles in solution as function of the scattering angle and concentration. The resulting K*c/R (with K…optical constant, c…concentration, R…Raleigh constant) for each angle is plotted and extrapolated to determine the molecular weight. The SEC

column separates the protein prior to the MALS measurement according to molecular weight to increase the signal quality (polydispersity). The S200 SEC column was equilibrated with running buffer containing DM and lipids until RI and scattering signal leveled at baseline. The scattering signals were calibrated with BSA.

200 µg of MaCCC E505Q622C were injected onto the column under oxidizing conditions. The chromatogram and the calculated masses are depicted in Fig 2-28. The elution volume of MaCCC of 11 ml was close to the low elution volume found in previous experiments from MaCCC in the presence of lipids. The corresponding molecular mass of the PDC (protein detergent complex) was 312 kD (0.2 % error). The

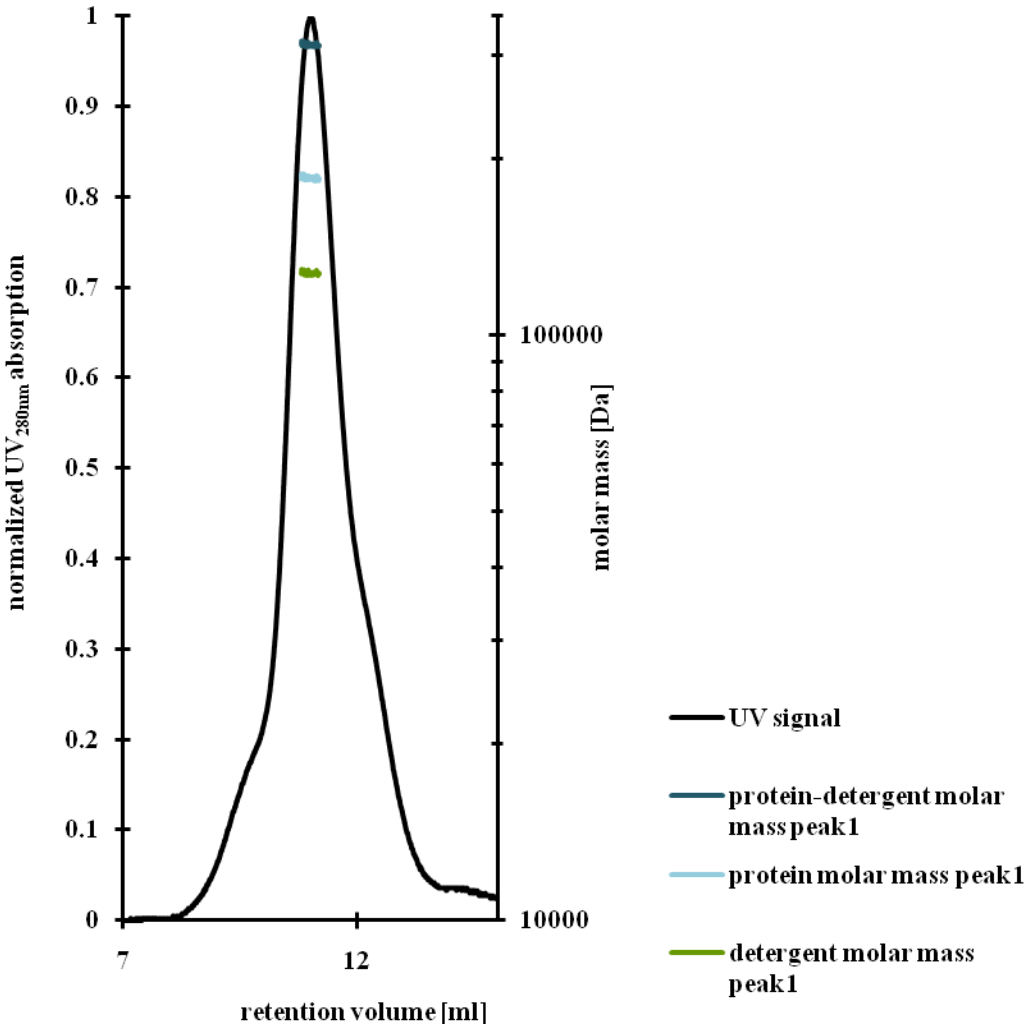


Fig 2-28 SEC-MALS of Oxidized MaCCC E505Q622C
The double mutant MaCCC E505Q622C containing cysteine mutations in the proposed dimer interface was analyzed using SEC-MALS with RI (refractive index) and UV detection under oxidizing conditions. The normalized UV_{280nm} absorption (left axis) and the molar mass (right axis) are plotted against the retention volume of an S200 column. The peak elution is 11.2 ml with the indicated mass of the PDC (protein detergent complex) (dark blue), protein core (light blue) and the detergent micelle around the protein (green).

weight of the protein core was calculated to be 185 kD (0.2 %) with a polydispersity M_w/M_n=1.000 (0.3 %) and the limits of 10.8 to 11.2 ml. Since the molar mass of a MaCCC monomer is 83 kD this peak thus most probably corresponds to the oligomeric state of a dimer.

When 200 µg of reduced MaCCC E505Q622C was subjected to the same analysis a different mass distribution was observed. Partial reduction of the cysteine bridges in the dimer interface led to formation of a double peak with retention volumes of 11.2 and 12.7 ml (Fig 2-29). The peak at 11.2 ml corresponds to the dimer observed before (PDC=300 kD (0.3%) with a protein core of 165 kD (0.3%), elution volume limits

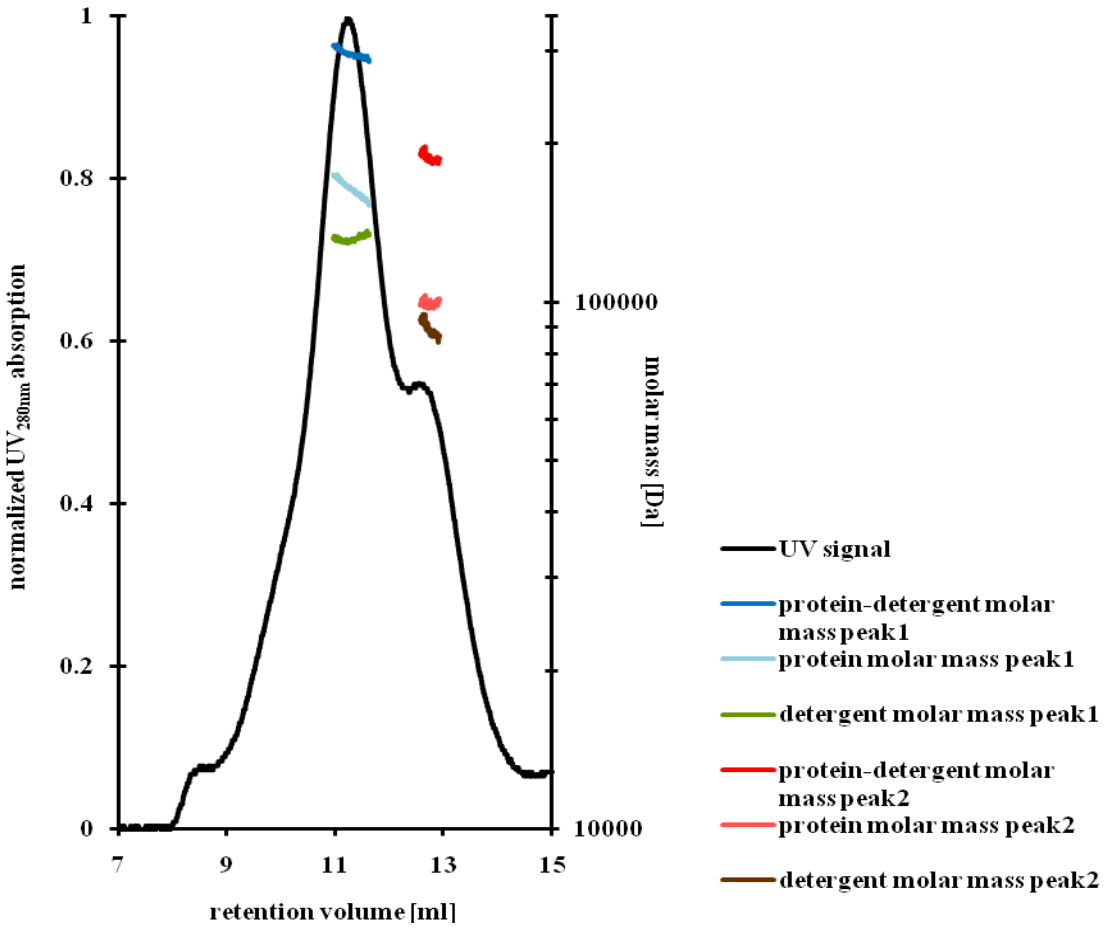


Fig 2-29 MALS of Reduced MaCCC E505Q622C
MaCCC with introduced cysteine mutations in the proposed dimer interface E505Q622C was analyzed using SEC-MALS with RI (refractive index) and UV detection after incubation of the protein with 0.5 mM TCEP for reduction of the cysteine bridges. The normalized UV_{280nm} absorption (left axis) and the molar mass (right axis) are plotted against the retention volume of an S200 column. The elution volume of the first peak is at 11.2 ml. The masses of the PDC (dark blue), protein core (light blue) and the detergent micelle around the protein (green) are indicated. The second peak belonging to the reduced protein has a retention volume of 12.7 ml. The molar masses of the PDC (red), protein core (pink) and the detergent micelle (brown) are indicated.

between 10.9 and 11.6 ml and $M_w/M_{n=}$ 1.001(0.5 %)). The second peak at 12.7 ml, however, belongs to a different oligomeric state. The mass of the calculated PDC was 188 kD (0.7 %), with a protein content of 99 kD (0.8 %), elution volume between 12.5 to 13 ml (M_w/Mn=1.000(1 %)).

The MALS experiment strongly suggests that the lower elution volume seen in SEC corresponds to a dimer of the protein, while the peak at the higher volume is due to MaCCC monomer. The formation of disulfide bridges confirms the proximity of the residues in the dimeric protein. These mutations can be used to stabilize the oligomeric state of MaCCC. Additionally, to fix MaCCC in a defined conformation on the transport cycle might prove to be essential for a future crystallization of the protein.

2.3.3.4 Summary

In the course of this project I have screened the overexpression and biochemical characterization of 17 prokaryotic homologues of cation chloride cotransporters. The broad screening efforts included the use of two different expression hosts, *E. coli* and *L. lactis*, and the fusion of N-terminal or C-terminal His-tags and C-terminal GFP fusions. Using *E. coli* and the strong T7 promoter system only one homologue, MaCCC, could be overexpressed in a DM/DDM soluble state (with an overall yield of 200µg of protein per liter of culture). The protein was purified and subjected to crystallization. Whereas other homologues, like SynCCC and SfCCC, were also overexpressed at suitable levels , they were only extractable with Fos-cholins and were prone to aggregation. Changing to the tuneable pBAD system facilitated the expression of other candidates such as SynCCC, PiCCC, PmaCCC and SrCCC in a putatively folded state (as judged by in-gel fluorescence of GFP fusion proteins). After extraction with DDM these proteins could be purified by IMAC and analyzed with SEC. During this procedure only MaCCC protein was identified as stable and monodisperse protein. Other well expressing homologues like SrCCC, PiCCC and PmaCCC showed strong aggregation behavior after solubilization. While SynCCC was purified as stable peak after gel filtration its yield was very low (50µg of protein per liter of culture). The expression of proteins carrying a N-terminal His-tag was unsuccessful for most constructs.

L. lactis offered the choice to express CCCs in a Gram-positive expression host. Expression under the control of the P_{NIS} promoter could be detected for most of the investigated homologues as C-terminal GFP and His-tag fusion proteins. With N-terminal His-tag fusion only SrCCC showed expression. PiCCC, SrCCC, and PmaCCC were purified in DDM and showed a symmetric elution peak on SEC column. The total yields of expressed proteins were so little that *L. lactis* was not investigated further and the priority was given to homologues expressing in *E. coli*.

2.3.4 Crystallization

All purified constructs were broadly screened for crystallization in a broad protein concentration range from 3-15 mg/ml. In the first crystallization trials with MaCCC

purified with His_6-tag only crystals from ubiquinol oxidase that co-purified as impurity were observed. With the improved protein purity and the supplement of lipids the crystallization was attempted in different detergents but still did not provide any promising leads.

Beside the wild-type protein, also the crosslinked MaCCC mutant with an introduced disulfide bond in the domain interface, which potentially stabilizes the protein, was subjected to crystallization screening but did not yet provide any promising results either.

2.3.5 Materials and Methods

Preparation of Genomic DNA (gDNA). The cell pellet of the prokaryotic organism was resuspended in 0.25 M Tris-HCl pH 8.0, 25 % (w/v) sucrose, 10 mg/ml lysozyme and incubated for 1 h at 37 °C. After addition of 1 % (w/v) SDS and 200 µg/ml proteinase K the lysis solution was again incubated for 1 h at 65 °C. The proteins were extracted once 1:1 with phenyl:chloroform:isoamyl acohol (25:24:1) pH 8.0 and the aqueous solution was washed at least three times 1:1 using chloroform:isoamyl alcohol (24:1). DNA was precipitatd by mixing the aqueous phase 1:1 with isopropanol and 5:1 with 7.5 M ammonium acetate and incubation for 30 min at -20 °C. The DNA was pelleted by centrifugation (15 min, 4 °C, 16000 g) and washed once with ice cold EtOH. The pellet was resupended in 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA pH 8.0 and kepped at 4 °C for immediate use and at -20 °C for storing.

PCR. For amplification of the gene different polymerases were used to obtain a full set of genes. The PCR was mixed according to the manual for Pfu polymerase (Fermentas), Vent polymerase (NEB) and thermoprime Mastermix (ABgene) supplemented with 50 µg of genomic DNA, 0.4 µM of forward and reverse primer and 4 % DMSO. The reaction was performed after hot start with 35 cycles of a temperature protocol that includes melting at 95 °C for 60 s, annealing for 45 s at 53 °C elongation according to the protocol of the polymerase. The PCR was finished by a final elongation step for 6 min at the optimal polymerase temperature. The PCR reaction mix was subsequent separated by agarose gel electrophoresis, and the expected DNA bands excised and purified with the NucleoSpin purification kit (Roche).

Ligation. For ligation the linearized plasmids was mixed with the digested insert at a molar ratio of 1:5 supplemented with reaction buffer and 400 U of T4 DNA ligase (NEB), adjusted to a final volume of 20 µl and incubated for 3 h at 16 °C.

Transformation of *E. coli*. 50 µl of chemically competent cells were thawed on ice and incubated for 10 min with the DNA. Heat shock was performed at 42 °C for 45 sec with following chilling on ice for another 10 min. After addition of 200 µl LB medium

the cells were grown for 1 h at 37 °C and plated on LB agar plates with incubation at 37 °C overnight until growth of colonies.

Expression Test of Prokaryotic Homologues. After transformation at least three colonies were picked for each construct and grown as precultures overnight at 37 °C in LB-medium. The preculture was diluted 100 times with the expression medium (mostly TB medium). After growing to an OD_{600nm}=0.5 at 37 °C the cultures were induced and grown for 3 h or they were cooled to 25 °C and induced with following incubation overnight. For analysis aliquots of roughly 1 mg of total protein (1ml of OD_{600nm}=3.3) were taken and the cell pellet prepared by centrifugation.

In-Gel GFP Fluorescence Detection from (Geertsma, Groeneveld et al. 2008). Whole-cell expression samples equivalent to ~1 mg of total protein were resuspended in 400 µl ice-cold lysis buffer (50 mM KP_i (pH 7.2), 1 mM $MgSO_4$, 10 % (w/v) glycerol, 1 mM PMSF and 20 µg/ml DNase). After addition of glass beads (300 mg, 0.1 mm diameter) the samples were broken in a FastPrep device (Bio101) for two times 20 s at force 6 with 5 min cooling on ice between the runs. 25 µg of total protein samples were analyzed by SDS/PAGE, and in-gel GFP fluorescence was visualized with an LAS-3000 imaging system (Fujifilm) and analyzed with the AIDA software (Raytest). Subsequently, gels were submitted to semidry electroblotting and immunodetection with. Chemiluminescence was detected using Fujifilm LAS-3000 imaging system.

Differential Sedimentation of Folded and Aggregated Proteins from (Geertsma, Groeneveld et al. 2008). Whole-cell expression samples were prepared as described for in-gel fluorescence. After bead beating, aliquots of 100 µl were solubilized for 60 min at 4 °C in the presence of 1 % (wt/vol) DDM with continous mixing. After saving a small aliquot (20 µl, 45 µg) to represent the sample before ultracentrifugation, the remaining sample was centrifuged at 355,000 × *g* for 15 min at 4 °C in a TLA 100.1 rotor. Both aliquots with 25 µg total protein were separeted by SDS-PAGE and analyzed by anti His-tag immunodetection.

Analysis of Overexpression with His-tag Detection. Whole-cell expression samples corresponding to 600 µg total protein were lysed in 100 µl of 5 M urea and 50 µg were separated by SDS-PAGE and analyzed by anti His-tag Western blot using a primary HRP coupled antibody against a His_6-tag (Roche)

Expression of Full-Length CCCs. Full-length MaCCC, transformed in *E. coli* BL21 (DE3) cells were grown as preculture overnight in LB medium. Cultivation started with 1 % (v/v) dilution into autoinduction media supplemented with 50 µg/ml kanamycin at 30 °C and incubation overnight. An onernight preculture of the additional CCCs

Isolation of Membrane Vesicles. Cells were harvested and lysed with an Emulsiflex C3 high pressure homogenizer (Avestin) in 50 mM Tris-HCl (pH 8.0), 150 mM NaCl (buffer A) with addition of 1 mg/ml lysozyme, 20 µg/ml DNase, 1 µg/ml leupeptin, 1 µg/ml pepstatin and 1 mM phenylmethyl sulphonyl fluoride (PMSF) with two passages with at least 700 bar. The lysate was cleared by low-spin centrifugation (30 min, 10000 x g) and the membranes pelleted with high-spin centrifugation (1 h, 100000 x g) and flash frozen in liquid N_2 and stored at -80 °C.

Purification of Prokaryotic CCC with His_{10}-Tag. The membranes were thawed on ice and resuspended in buffer A (50 mM Tris-HCl pH 8.0, 150 mm NaCl) + 10 % (v/v) glycerol at a ratio of 10 ml buffer A per 1 g of membranes. Solubilization was started with the addition of 2 % [w/v] DM (or other detergent) and incubation for 2 h with subsequent centrifugation (30 min, 100000 x g). The supernatant was incubated in batch with equilibrated Ni-NTA (Quiagen) after addition of 15 mM imidazole for 30 min. The resin was washed with buffer A + 50 mM imidazole on column until baseline and the protein eluted with buffer A + 300 mM imidazole. The elution fraction was immediately supplemented with 5 mM EDTA to complex leaking Ni^{2+}. For cutting of the His-tag HRV 3C protease (GE Healthcare) was added at a molar ration 1 mol protease for 5 mol target protein. After incubation for 2 h the protein solution was concentrated to 500 µl with an Amicon Ultra Centrifugal filter device (MWCO=100 kD, Milipore). The concentrated protein was injected onto a gel filtration column (Superdex 200 10/300 GL, Äkta Prime system, GE Healthcare) equilibrated with running buffer (10 mM Tris-HCl pH8.0, 150 mM NaCl, 5 mM DM). Elution was recorded with absorption at 280 nm. All steps were carried out at 4 °C.

Detergent Exchange. After SEC the new detergent was added to the protein at a concentration of 3xCMC and dialyzed against 500 x sample volume dialysis buffer with 3xCMC of the new detergent overnight at 4 °C. For testing the stability in the new detergent an additional SEC was done using the new detergent.

Crystallization. After SEC the protein was concentrated with an Amicon Ultra Centrifugal filter device (MWCO=100 kD, Milipore) to 5-12mg/ml. In a standard sitting drop experiment 100nl of protein solution were mixed with 100nl precipitant solution by a robotic system at 20 °C and incubated against the precipitant solution at 4 °C and 20 °C

Mutation and Disulfide Bridge Formation. Mutations were inserted with the QuikChange method (Stratagene) and confirmed by sequencing. Mutant proteins were purified following the same protocol as for WT. For crosslinking studies all constructs had the His-tag removed. The respective disulfide bridges formed spontaneously upon

exposure to air or after addition of 1 mM $Na_2S_4O_6$ and were reduced by addition of 5 mM β-mercaptoethanol or 5 mM TCEP. The proteins were analyzed by SDS-PAGE.

Multi-Angle Light Scattering. MALS was performed with an experimental set up consisting of an Äkta explorer with a S200 column (GE Healthcare) and UV-detector (A900, GE Healthcare) followed by a differential refractometer (OPTILAB rEX, Wyatt technology) and a miniDAWN MALS detector (Wyatt technology). The columns and detectors were equilibrated with running buffer (10 mM phosphate buffer pH 7.5, 150 mM NaCl, 5 mM DDM,) until baseline and a calibration run was performed with 100 µg BSA to align the signals. 100 µg of purified protein was injected onto the column and the UV_{280nm}, the refractive index (RI) and the scattering signals were recorded. The data was analyzed using Astra V software (Wyatt technology). Since the protein accounts with known amount to the UV absorption and RI and the detergent DDM only to the RI the protein contribution to the protein detergent complex MALS signal can be calculated. The modifier dn/dc for the detergent was set to 0.47 based on oral communications with T. Huber (Dissertation Thomas Michael Huber 2007, University of Zurich).

2.4 C-terminal Domain of CCCs

Soluble domains of transport proteins frequently play an important role in the regulation of transport in response to phosphorylation or ligand binding. They can serve as interaction platform with other proteins. These domains can often be expressed as soluble proteins, which makes them easier amendable to crystallization and structure determination. Additionally the isolated domains facilitate biochemical investigations such as binding assays to identify regulating ligands. These experiments are easier on soluble domains compared to labile, detergent solubilized membrane proteins. Moreover, the structure solution of these domains provides initial insight into the organization of the protein which is valuable for the purification, crystallization and structure solution of the full-length transporters. In cation chloride cotransporters the C-terminal domain accounts for 30-40% of the protein and is predicted to form a folded structure. While for eukaryotic CCCs the domain is a target for phosphorylation, so far no ligand was found to bind to specific sites in the proteins.

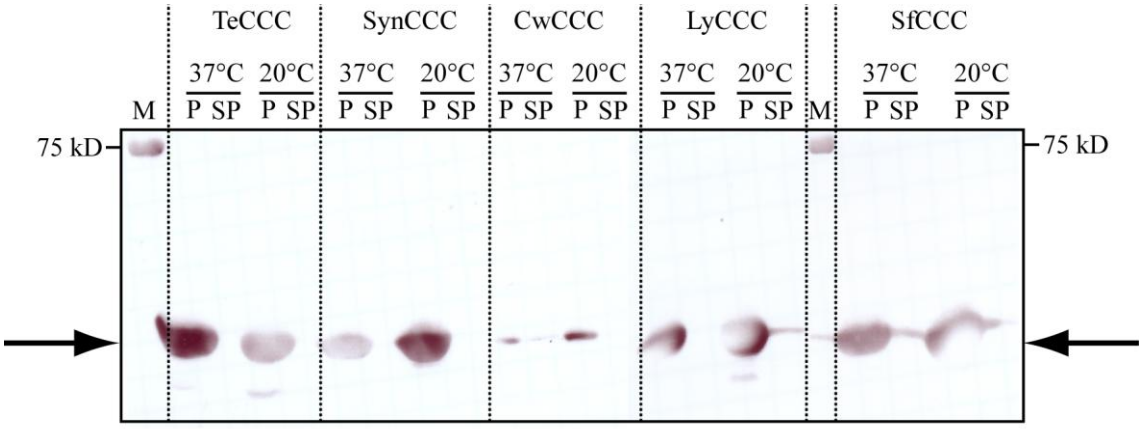


Fig 2-30 Expression Test of Prokaryotic C-Terminal Domains
The C-terminal domains of the depicted homologues were expressed in LB medium using *E. coli* BL21 (DE3) cells. After induction at OD_{600nm}=0.5 the cells were incubated for 3 h at 37 °C or overnight at 20 °C. The samples were centrifuged following cell lysis and analyzed by Western blot. The signal corresponding to the domains is indicated by an arrow. P... pellet, SP... supernatant

Different constructs of the isolated domain of prokaryotic CCCs were designed on the basis of hydrophobicity predictions that indicate the boundary between the membrane part and the start of the domain. After the successful structure determination of the domain of MaCCC the expression constructs of the cytoplasmic domains of other homologues were structure-based. The expression of the C-terminal domains from different prokaryotic homologues proved to be very difficult, since all C-terminal domains (Cdom) besides the MaCCC Cdom were expressed in inclusion bodies. The variation of temperature, media, inducer concentration and construct length had no influence on this behavior. Also the fusion of a solubility tag like MBP lacking the export signal did not change the excessive aggregation of the protein in the cell. Multiple constructs were designed based on stepwise truncations of 5 amino acids. The C-terminal domains of different homologues (Tab 2-1) were expressed in *E. coli* BL21 (DE3) cells in LB medium. Expression was induced at OD_{600nm}=0.5 with 0.5 mM IPTG for 3 h at 37 °C or overnight at 20 °C. The cells were lysed and the soluble fraction separated by centrifugation (Fig 2-30). Apart from the MaCCC C-terminus, all investigated constructs sedimented through centrifugation, indicating that they were expressed in inclusion bodies. Only the C-terminal domain of MaCCC allowed soluble expression and was used for functional investigations and structure determination.

2.4.1 *Methanosarcina acetivorans* C-terminal Domain

The C-terminal domain of MaCCC is 32kD large and constitutes nearly 40% of the full-length protein. The first construct used for expression trials of the C-terminal domain from MaCCC was designed to start approximately after the predicted last transmembrane helix at residue H450 (Fig 2-31).

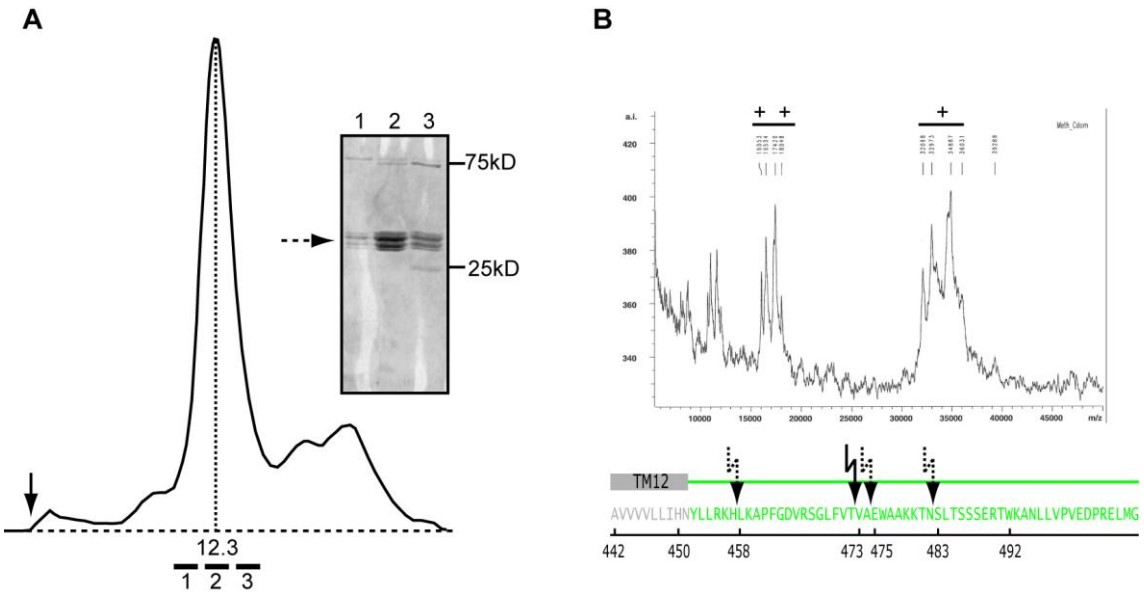


Fig 2-31 MaCCC C-Terminal Domain Construct Design
The truncation of MaCCC to have the C-terminal expression construct was initially based on hydrophobicity analysis. The last transmembrane segment is assigned [TM12, gray] as well as the cytosolic part belonging to the C-terminal domain [green]. The first construct started at H450. This construct showed several degradation products on the protein level analysed by MS at its N-terminal region (dashed flashes). Combined with the experience of purification of full length MaCCC which also fell apart (flash), a new construct was designed starting at residue V473 which was highly expressed and stable.

The domain was initially expressed with a C-terminal His_6-tag fusion without protease site for 3 h at 37 °C using *E. coli* BL21 (DE3) cells and 0.5 mM IPTG. These conditions gave sufficient protein for initial purification trials. The IMAC purification started with binding to a Ni-NTA column and washing with 15 mM imidazole. The protein was completely eluted from the resin with 300 mM imidazole. SEC revealed a predominant peak at 12.3 ml corresponding to an apparent molecular weight of 175 kD. SDS-PAGE analysis showed that the protein is partly degraded (Fig 2-31A) and allowed the identification of four degradation products. Since anti His-tag Western blot analysis showed a signal for all 4 protein fragments, the degradation occurred at the N-terminus which left the C-terminal His_6-tag attached. MALDI-MS analysis identified three specific breaking points at the N-terminal tail, L458, E475 and S483 (Fig 2-31B, dashed flashes). The low yield of only 50 µg purified protein per liter of expression culture also indicated a suboptimal expression construct. Partial proteolysis in the linker region separating the membrane part and the C-terminal domain was already observed during purification of the full-length transporter. In this case the cleavage was identified by MALDI-MS at residue V473 (Fig 2-31B, solid flash). The new construct was thus designed combining the information from the first domain purification trials and the experience from the purification of full length MaCCC. With a construct starting at residue V473 a clean, monodisperse, stable protein species could be purified with a

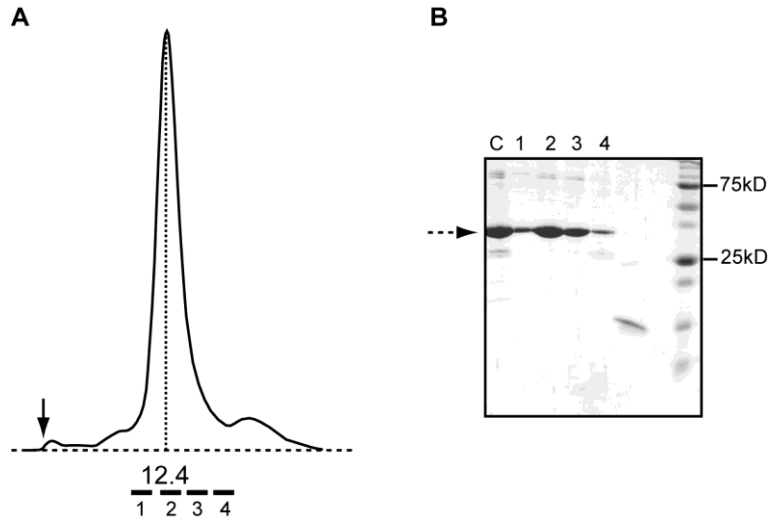


Fig 2-32 MaCCC Cdom Purification
(A) After redesign of the expression construct, MaCCC Cdom was purified and eluted as single monodispers peak on SEC at 12.4 ml. Nearly no aggregates were present (arrow). The four main peak fraction (1-4) were further analyzed on SDS-PAGE. The retention volume indicated in [ml] is plotted against the absorption at 280 nm. **(B)** Coomassie stain of the main peak fractions (1,2,3,4) and the concentrate (C) of fraction 2 and 3.

yield of 3 mg purified protein per liter of expression culture (Fig 2-32). The elution volume corresponded to an apparent molecular weight of 170 kD which would be in the appropriate range of either a pentamer (5*33 kD=165 kD) or a hexamer (6*33 kD=198 kD).

Based on the truncated domain construct a second construct was designed that includes a C-terminal 3C protease site to remove the His_6-tag. Both protein constructs were used for the characterization by X-ray crystallography and analytical ultracentrifugation (AUC).

I have used AUC to get quantitative insight into the oligomeric organization of the MaCCC Cdom. AUC measures the influence of a centrifugal field on the concentration distribution of the traced compound. The experiment can be set up as sedimentation-velocity or sedimentation-equilibrium experiment. Sedimentation-velocity AUC allows the determination of the sedimentation coefficient S and approximate estimation of the molecular weight (Schuck 2000). Sedimentation-equilibrium AUC in contrast allows the measurement of the diffusion coefficient, which can be used to calculate the exact molecular mass, with the determined S of the velocity experiment. The measured S is still dependent on the buffer composition, temperature and the protein concentration. The calculation of S in water at 20 °C at infinite dilution abolishes the influence of measurement parameters and creates a condition independent $S_{20, H2O}$. The sedimentaion coefficient allows only indirect conclusion to the molecular weight.

All measurements described here were carried out as sedimentation-velocity experiments at 15 µM protein concentration, pH=7.0 and 4 °C (Fig 2-33). The MaCCC Cdom lacking the protease site and still containing the His_6-tag sediments at a S=9.0 (blue curve). In contrast, the removal of the His_6-tag, or the increase of the pH to 9.5 led

to a shift to a smaller sedimentation coefficient at S=2.4 (red curve). Finally, the mutation of an essential interface residue responsible for major interactions in the interaction interface shifted S to 1.7 (for discussion see Publication 2.4.1.2). By converting the sedimentation coefficient to its $S_{20, H2O}$ value and by taking the diffusion coefficient and molecular shape of the protein into consideration, these sedimentation behaviors correspond to the molecular masses of 9.0 S=190 kD, 2.7 S=51 kD and 1.7 S=30 kD. Since the mass of a subunit of MaCCC Cdom is 33 kD, the peaks correspond to a hexamer, dimer and monomer, respectively.

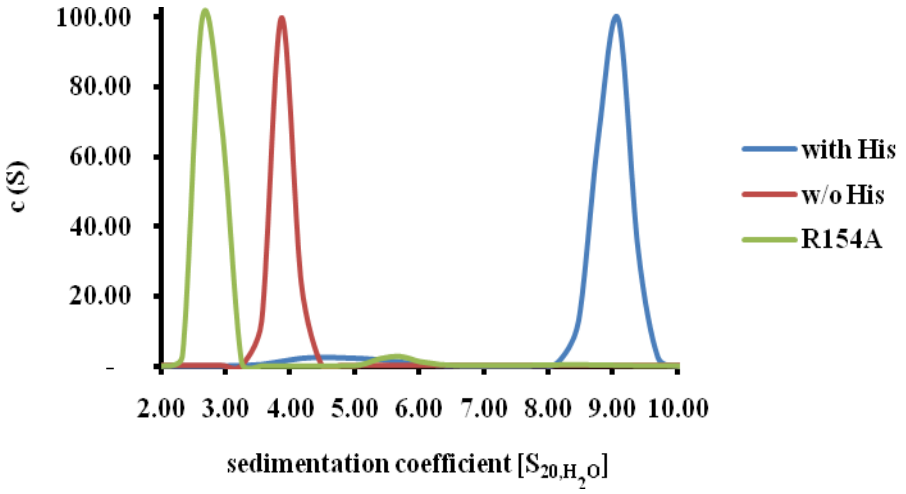


Fig 2-33 Analytical Ultracentrifugation of MaCCC Cdom
The distribution of the sedimentation coefficient (c(S)) of an AUC experiment with different constructs of MaCCC Cdom is plotted against the sedimentation coefficient. MaCCC Cdom His_6-tag (blue, with His) sediments at 9 S. Introduction of a 3C protease site and removal of the His-tag resulted in a shift to S=2.7 (red, w/o His). The mutation of an essential interface residue R154A resulted in S=1.7 (green, R154A).

For structural analysis MaCCC Cdom His_6 was subjected to crystallization screening. In a sitting drop experiment crystal were observed at a condition containing 100 mM MES (pH 6.5) and 1.1 M $MgSO_4$. These crystals belong to the trigonal space group $P3_221$ and diffracted to a resolution of 1.9 Å and allowed the collection of a complete data set.

To determine the phases of the structure factors the protein was expressed in minimal medium supplemented with all amino acid besides methionine. Seleno methionine (Se-

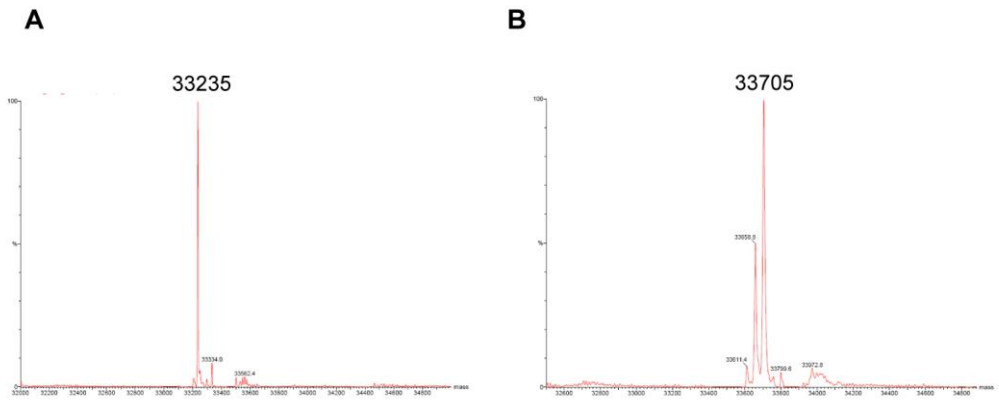


Fig 2-34 MaCCC Cdom Se-Methionine Derivatization
ESI-TOF-MS of **(A)** native MaCCC Cdom and **(B)** Se-Methionine derivatized MaCCC Cdom. The relative intensity (highest peak=100%) is plotted against mass. The mass of the peaks with the highest intensity is labeled.

Met) was added to be incorporated at methionine positions and the incorporation was confirmed by mass spectrometry (Fig 2-34). Since MaCCC Cdom has 10 methionines in its sequence, the replacement by Se-Met results in a mass difference of 470 Da in case of complete derivatization, which was observed in the MALDI-TOF-MS (Se-Met protein 33705 Da, wild-type protein 33235 Da, Δ=470 Da).

Crystals of Se-Met protein grew in the same condition as the native protein. A

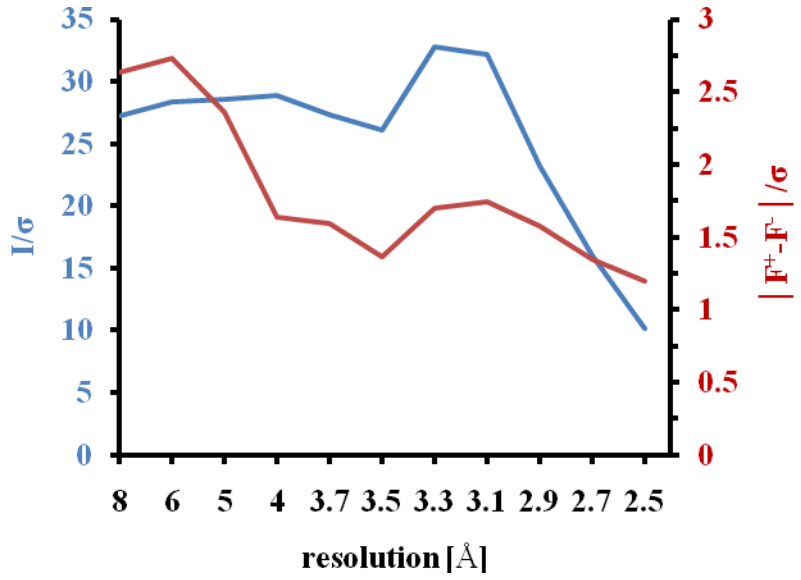


Fig 2-35 MaCCC Cdom SAD Data Collection Statistics
After collection of Se methionine substituted crystals the collection parameters were plotted against resolution. Intensity over standard deviation (I/σ, blue, left axis), anomalous difference over standard deviation ($|F^{+}-F^{-}|/\sigma$, red, right axis).

dataset of these crystals was collected at the selenium κ absorption edge for maximizing the anomalous differences. The data collection was performed at high redundancy to improve the statistics of the anomalous signal. After scaling and merging of the data (where Friedel pairs were kept separated) the magnitude of the anomalous contributions were analyzed by plotting the intensity (I/σ) and the anomalous difference ($|F^{+}-F^{-}|/\sigma$) against resolution (Fig 2-35). At a resolution of 2.5 Å the anomalous difference was still above background (1.2) which was chosen as cut-off for phase determination.

The structure was determined by single wavelength anomalous dispersion (SAD). Since the asymmetric unit contains a single subunit, as indicated by calculation of the Matthews coefficient, 9 selenium sites were expected to contribute to the anomalous scattering assuming the first Se-Met to be unordered. Using the programs SHELXC and SHELXD 7 selenium sites were located and refined (Fig 2-36).

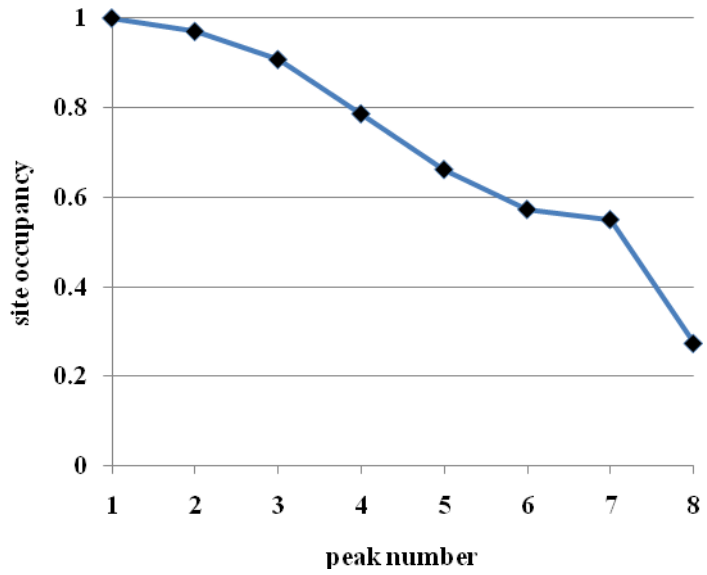


Fig 2-36 Heavy Atom site Occupancy After Heavy Atom Search in SHELXD
The heavy atom site occupancy is plotted against the Se sites. 8 sites could be determined including 7 with high occupancy (1-7).

Initial SAD phases were improved by 20 cycles of density modification and solvent flipping with SHELXE, which allowed the determination of the correct hand of the selenium coordinates (Fig 2-38).

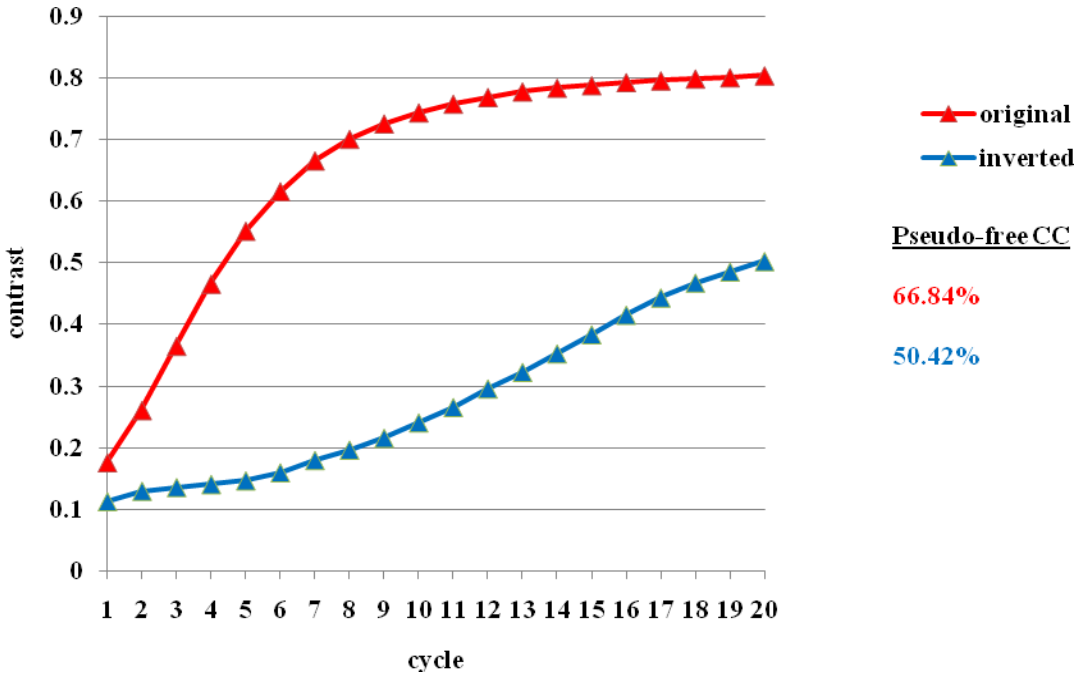


Fig 2-38 Density Modification in SHELXE
Progress of density modification. The contrast is defined as the variance of V (variance of the density on spherical surface with radius 2.42 Å) averaged over all pixels of the electron density. The pseudo-free correlation coefficient (Pseudo-free CC) is calculated at the end by leaving out 10 % of reflection at random and doing one round of density modification with application of the improved phases. If the density modification converges and the difference between both solutions is high, the right enamtiomorph can be distinguished. The original solution (red) strongly improved the contrast and converged. In combination with the high Pseudo-free CC these were good indicators for the right solution.

The improved electron density allowed model building and initial refinement in

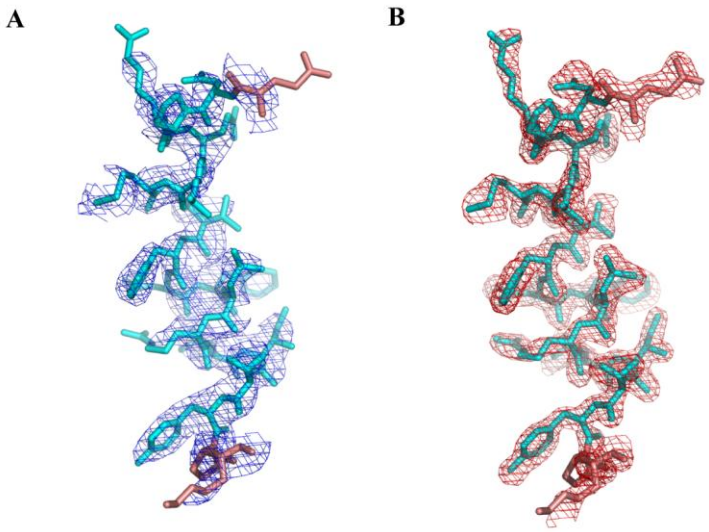


Fig 2-37 Electron Density of MaCCC C-Terminal Domain
(A) Initial experimental electron density and **(B)** 2Fo-Fc electron density after completed refinement. The density was calculated to 2.5 Å and 1.9 Å for A and B respectively and contoured at 1.5 σ. The electron density is superimposed on the refined structure of MaCCC Cdom (residue 28-46 with α helix residues, light blue and loop residues, light red). The experimental map was calculated from native amplitudes and solvent flattened SAD phases.

Phenix. The final refinement was carried out against the native data set at 1.9 Å by simulated annealing with Phenix. The improvement of the electron density during refinement is illustrated in Fig 2-37. Compared to the initial density calculated from experimental phases the final model phases improved considerable. While the initial electron density clearly showed the secondary structure elements and side chain conformations of bulky residues like tyrosines, the final density additional unambiguously defines the atomic positions of the main and side-chain atoms.

2.4.1.1 Additional Crystal forms

Upon the broad crystallization screening of the MaCCC C-terminal domain different crystal forms were identified that show variations in the crystal packing and thus provide insight into the oligomeric organizations of the domains. The first crystal form observed that also allowed the high resolution structure determination to 1.9 Å was of

construct	with - w/o His tag	crystal form	resolution [Å]
domain-His_6	with	$P3_221$	1.9
domain-3C-His_6	w/o	C222	3.0
domain-3C-His_6	w/o	$I4_1$	2.85

Tab 2-6 Crystal Forms of MaCCC C-Terminal Domain

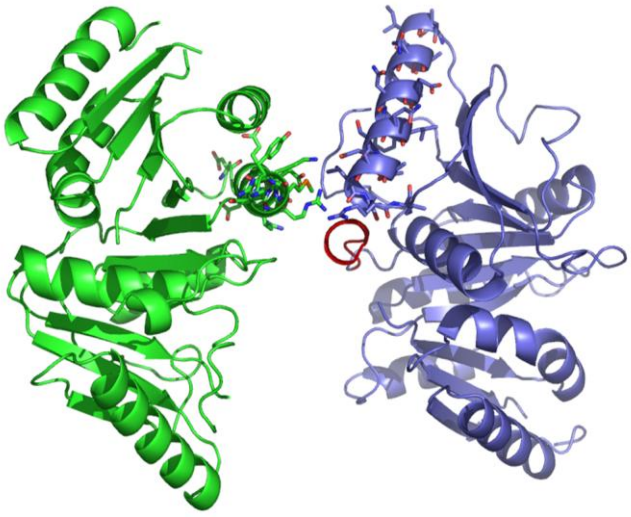


Fig 2-39 Crystal Contact in Crystal Form $P3_221$
Two neighbouring domains in crystal form $P3_221$ are depicted in blue and green. The residues of the interacting helices are shown as sticks. The C-terminus of the blue domain (red) forms a short α-helix. Additional residues with α-helical propensities elongate this helix that can interfere with the crystal contact.

space group $P3_221$. This construct had the His_6-tag attached directly after the protein, which was not removed upon crystallization. In an attempt to observe different crystal packing, to distinguish between native interactions and crystal contacts and thus to define the native dimer interface the construct was varied. A protease site between the protein and the His_6-tag was introduced, which allowed the crystallization and structure determination of the MaCCC C-terminal domain in the space groups C222 and $I4_1$ by molecular replacement.

Although the crystal forms belong to different space groups important features of the crystal contacts observes in the $P3_221$ space group are also observed in other crystal forms.

In Tab 2-6 it is illustrated that the length of the C-terminus and the presence of the His-tag had a large influence on the crystal symmetry and resolution. The analysis of the packing requires a thorough inspection of the crystal contacts. One contact is mediated by few residues near the C-terminus that protrudes from the surface (Fig 2-39). For the high resolution structure of the protein all residues were resolved including the first histidine of the His_6-tag. The last five histidines of the tag were disordered and thus not resolved in the structure. In this case six additional amino acids, histidines, were added to the C-terminus. For the constructs harboring a protease site 14 amino acids are added to C-terminus that reduces to six additional residues upon cleavage of the tag. These six residues (LEVLFQ) are remaining on the terminus (Fig 2-14) and they have large α-helical propensities and could thus interfere with a proper packing of the crystals.

Protease	**Cleavage Site**	**C-Terminal Overhang [bp]**
TEV	E-Xaa-Xaa-Y-Xaa-Q-*(G/S)	6
HRV 3C protease	L-E-V-L-F-Q-*G-P	6
Factor Xa	I-(E/D)-G-R*	4
Thrombin	L-V-P-R-*G-S	4
Enterokinase	D-D-D-D-K*	5

Tab 2-7 Protease Sites and Their Overhangs

The introduction of additional residues at the termini of the protein has to be considerably investigated since these might influence crystallization. The protease sites most commonly used for crystallization are suited for N-terminal cleavage because in this case they only add 2-3 residues to the protein. If they are used C-terminally they

add up to six residues (Tab 2-7). Only Factor Xa and thrombin cleavage sites produce comparably short overhang. The disadvantage of these proteases however, is the low catalytic activity at 4 °C where purifications of moderately stable proteins have to be carried out.

The statistics for data collection and structure determination and a detailed analysis of the crystal packing is described in the following chapter as part of a published manuscript.

2.4.1.2 Publication

Structure
Article

X-ray Structure of the C-Terminal Domain of a Prokaryotic Cation-Chloride Cotransporter

Stefan Warmuth,[1] Iwan Zimmermann,[1] and Raimund Dutzler[1,*]
[1]Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
*Correspondence: dutzler@bioc.uzh.ch
DOI 10.1016/j.str.2009.02.009

SUMMARY

The cation-chloride cotransporters (CCCs) mediate the electroneutral transport of chloride in dependence of sodium and potassium. The proteins share a conserved structural scaffold that consists of a transmembrane transport domain followed by a cytoplasmic regulatory domain. We have determined the X-ray structure of the C-terminal domain of the archaea *Methanosarcina acetivorans*. The structure shows a novel fold of a regulatory domain that is distantly related to universal stress proteins. The protein forms dimers in solution, which is consistent with the proposed dimeric organization of eukaryotic CCC transporters. The dimer interface observed in different crystal forms is unusual because the buried area is relatively small and hydrophilic. By using a biochemical approach we show that this interaction is preserved in solution and in the context of the full-length transporter. Our studies reveal structural insight into the CCC family and establish the oligomeric organization of this important class of transport proteins.

INTRODUCTION

The cation-chloride cotransporters (CCC) constitute a large and ubiquitous family of secondary active ion transport proteins that mediate the coupled flow of Cl^- and the cations Na^+ or K^+ across the plasma membrane. In humans the family encompasses nine members that are grouped in different subbranches depending on the nature of the cation that balances the negative charge of Cl^- during electroneutral cotransport (Gamba, 2005). Although the two sodium-potassium-chloride cotransporters (NKCC1 and NKCC2) and the sodium-chloride cotransporter (NCC) mediate the uptake of Cl^- into the cell powered by the energy stored in the electrochemical potential of Na^+, the four potassium-chloride cotransporters (KCC1–KCC4) use the transmembrane K^+ gradient to extrude Cl^- from the cytoplasm into the extracellular medium. Two orphan members have not been assigned to any of the three subgroups (Gamba, 2005). The family members are involved in a variety of physiological processes ranging from the transport across epithelia of the kidney to the control of the cell volume and the regulation of electrical excitability in neurons (Gamba, 2005; Payne et al., 2003). Their importance is also emphasized by mutations in NKCCs and NCCs that cause kidney diseases such as Bartter's and Gitelman's syndrome, and by the fact that malfunction in KCCs has been connected to neuropathies as epilepsy and trauma (Hebert et al., 2004; Payne et al., 2003). NKCCs and NCCs are the receptors for loop diuretics, a major family of drugs involved in the treatment of hypertension, whereas KCCs have been proposed as potential targets for the treatment of epilepsy and sickle cell anemia (Hebert et al., 2004).

Despite the differences in the selectivity of the transported ions and the direction of transport in a cellular context, all family members share a common structural framework that consists of a highly conserved transmembrane domain that is framed by hydrophilic regions on its N and C terminus that are both located in the cytoplasm (Figure 1A) (Isenring and Forbush, 2001; Park and Saier, 1996). Functional investigations in different NKCCs have located conserved sequence stretches in the transmembrane domain that play a critical role for ion binding and transport (Isenring et al., 1998a, 1998b). The extended cytoplasmic N terminus is poorly conserved, of variable length, and has a large degree of predicted disorder, and might thus be unstructured (see Figure S1 available online). The C terminus, in contrast, is predicted to fold into a compact structure. The moderate sequence conservation of this region, although it is considerably lower than in the transmembrane domain, suggests that its structure is conserved throughout the family. Cytoplasmic components that are attached to the transmembrane catalytic domain are frequently found in transmembrane transport proteins and have in several families been shown to play an important role in the regulation of transport in response to cellular signals (Barabote et al., 2006). Although the role of the cytoplasmic domains of CCC transporters is still poorly understood, both regions have been implicated to be involved in regulatory processes. For example, the N terminus in NKCC1 carries recognition sites for kinases that were shown to modulate transport activity upon phosphorylation (Dowd and Forbush, 2003; Gimenez and Forbush, 2005), and the C terminus of KCCs was found to be involved in the activation of the protein in response to cell swelling (Mercado et al., 2006).

Although CCC transporters were originally discovered in animals, the recent abundance of genomic sequence information has allowed the identification of prokaryotic homologs of the family. Although smaller than their eukaryotic counterparts, the proteins share the most important structural features and probably also the basic mechanisms of transport. As among eukaryotic transporters, the sequence similarity is strongest in the transmembrane domain where several regions of functional importance are conserved (Figure 1B). The mismatch in size is due to insertions at both termini and in presumed loop regions

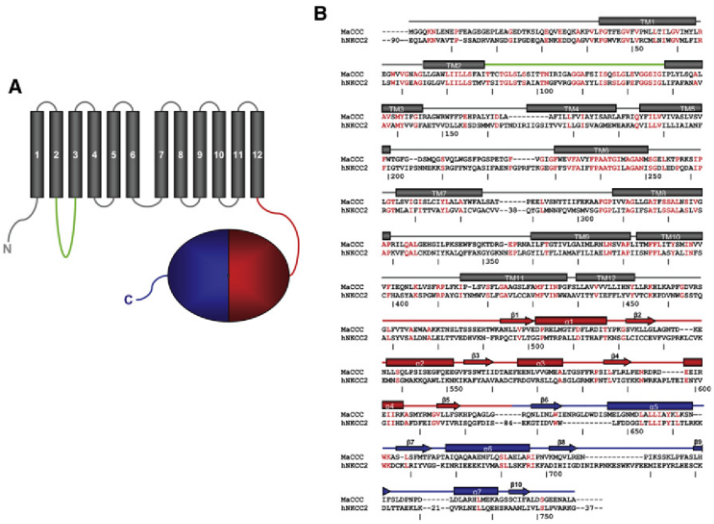


Figure 1. Conservation of CCC Transporters

(A) Putative topology of a CCC transporter. The 12 hydrophobic regions attributed to potential transmembrane spanning segments are represented in gray. The putative loop region that bridges segments two and three that has been shown to be critical for ion binding is colored in green. The two subdomains of the C-terminal cytoplasmic domains are colored in red and blue, respectively. (B) Pairwise sequence alignment of MaCCC (MA4506) and human NKCC2 (SLC12A). The alignment was made with ClustalW (Thompson et al., 1994) followed by manual adjustments in the region of the cytoplasmic domain (in accordance with the alignment shown in Figure S2). Identical residues are colored in red. The numbering corresponds to MaCCC. Elements of secondary structure are indicated above (hydrophobic stretches of the transmembrane domains are shown as helices and colored in gray, and the two halves of the cytoplasmic domain are colored in red and blue, respectively).

in eukaryotic transporters. Several of these inserted sequences are predicted to be unstructured (Figure S1). The long cytoplasmic N terminus is missing in prokaryotic transporters, whereas the C-terminal domain, although shorter, is essentially preserved. Similar to eukaryotic CCC transporters, the conservation in the C-terminal domain is lower than in the transmembrane part. However, a careful sequence alignment between different pro- and eukaryotic transporters suggests that its overall organization is conserved (Figures 1B and S2).

Although the transport properties and pharmacology of different members of the CCC family have been the subject of investigations during the last few years (Gamba, 2005), our knowledge of the structural organization of these proteins is still scarce. Different biochemical studies employing chemical crosslinking have suggested that the transporters form homodimers and that the C-terminal domain plays an important role in dimer stability (Moore-Hoon and Turner, 2000; Parvin et al., 2007). Apart from their oligomeric organization, the structural constituents of the family are unknown.

Here we present the structure of the isolated cytoplasmic domain of a prokaryotic CCC transporter from the archaea *Methanosarcina acetivorans* (MaCCC). The structure has an α/β fold that is distantly related to universal stress proteins. The protein forms dimers in solution and different crystal forms disclose a dimeric organization with an unusual interface that is relatively small and hydrophilic. In an attempt to probe whether this interface is conserved in solution, we have disrupted dimerization by removing critical interactions by mutagenesis and we have introduced cystein residues in adjacent positions to form specific intersubunit crosslinks. By using the same mutations to crosslink the full-length transporter, we show that the interactions are preserved in the context of the transmembrane protein. Our studies thus reveal first structural insight into the cytoplasmic domains of CCC transporter and establish an oligomeric organization that is likely general for the family.

RESULTS

Domain Structure

The cation-chloride cotransporter MaCCC is a 758-residue-long protein that is closely related to its mammalian counterparts. The largest homology is to the branch that includes the sodium-dependent transporters NKCC1, NKCC2, and NCC. MaCCC shares 24% of identical residues in a pairwise sequence alignment with human transporter NKCC2 (Figure 1B). The sequence conservation in the transmembrane region (31% of identical residues) is significantly higher than in the cytoplasmic domain (12% of identical residues). Similar to other bacterial CCC transporters, the long cytoplasmic N-terminal region is absent. In an attempt to reveal the structural organization of the cytoplasmic components of CCC transporters, we have cloned the isolated C-terminal domain of MaCCC. The domain was expressed as soluble protein, and was purified and crystallized. The structure was determined by X-ray crystallography at 1.9 Å resolution with phases obtained by the selenomethionine (SeMet) single-wavelength anomalous dispersion (SAD) method (Figures 2A and S3). This crystal form is of space group $P3_221$ with one copy of the domain in the asymmetric unit. The cytoplasmic domain of MaCCC contains 288 amino acids and forms a compact trapezoid-shaped elongated structure with overall dimensions of 30 × 40 × 60 Å (Figures 2B–2D). The protein has a mixed α/β fold and consists of two structurally related subdomains. Each subdomain contains a central five-stranded parallel β sheet. The β strands are, with exception of a loop connecting β8 and β9 in the second subdomain, bridged by α helices that pack on either face of the sheet. One helix, α1 in the first subdomain, is kinked halfway at the position of a glycine residue. Both subdomains are related by a pseudo two-fold arrangement with an axis located in the center of the two terminal β strands. The resulting antiparallel orientation of β5 and β10 connects the two elements of secondary structure to a continuous ten-stranded twisted

Structure
CCC Domain Structure

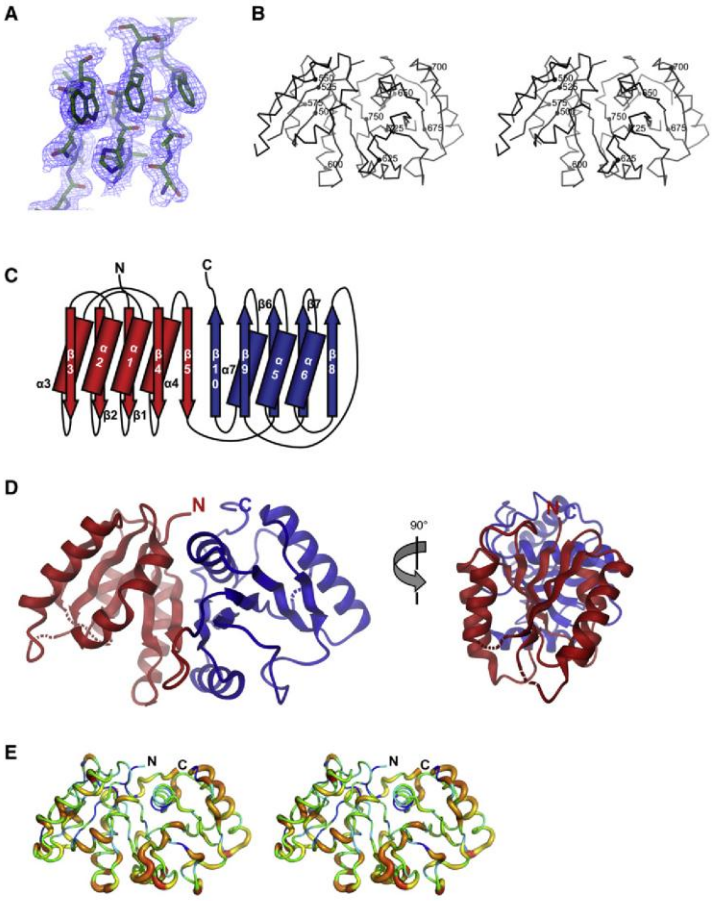


Figure 2. Structure of the MaCCC Domain Subunit

(A) $2F_o$-F_c electron density of a representative region of the MaCCC domain. The electron density (calculated at 1.9 Å resolution and contoured at 1σ) is shown superimposed on selected residues.

(B) Stereoview of a C-α trace of the MaCCC domain. Selected residues are labeled according to their position in the MaCCC sequence.

(C) Topology of the MaCCC domain. The two subdomains are colored in red and blue respectively.

(D) Ribbon representation of the MaCCC domain in two orientations. The relationship between the two views is indicated. The color coding is as in C.

(E) Stereo representation of a trace of the MaCCC backbone. The thickness and color coding represents the conservation of the respective region in a multiple sequence alignment of the domains from different organisms as shown in Figure S2. The range is from highly conserved residues displayed as thin blue tubes to poorly conserved regions displayed as thick orange tubes.

Figures 2–4 were prepared with DINO (www.dino3d.org).

β sheet. The extensive contact surface between the two subdomains encompasses 2790 $Å^2$ and includes, next to the central two-stranded antiparallel β sheet, interactions between residues in helices and loop regions. This domain structure shows resemblance to dimers of bacterial universal stress proteins (USP) that have been implicated to play an important role in signaling in bacteria and plants, although there is no relationship on the level of sequence (Figure S4) (Kvint et al., 2003).

Despite the moderate conservation in the cytoplasmic region, there is evidence that the structure of the MaCCC domain is preserved throughout the family. A sequence comparison between different prokaryotic domains and the domain of human NKCC2 aligns the elements of secondary structure observed in MaCCC and puts long insertions of the human protein into loop regions (Figures 1B and S2). This alignment also places the regions of largest conservation into the hydrophobic core, as expected for proteins that share the same fold (Figure 2E). The analysis of the sequence of the human protein with respect to its probability to fold into a defined structure reveals that several of the long insertions are likely unstructured (Figure S1). This distribution of long unfolded regions inserted between elements of secondary structure might thus account for the fact that, despite the overall conservation of its structure, the size of the domains ranges from 288 to over 450 residues.

Oligomeric Organization

The investigation of the sedimentation properties of the MaCCC domain in solution by analytical ultracentrifugation reveals the presence of higher oligomers. Proteins containing a His_6-tag (which was used for purification) sediment with an apparent molecular weight corresponding to a hexamer at high pH (Figure S5). Upon lowering the pH, the hexamers dissociate into smaller units that sediment as dimers. A dimeric organization of the protein in a broad pH range is also observed after proteolytic cleavage of the His-tag at a specific site at the C terminus of the domain. The dimeric organization is consistent with crosslinking studies on different human family members that were suggested to form homodimers. Because the oligomeric organization of the family appears to be preserved in the structures of the MaCCC domain, it is obviously interesting to see whether this organization would be revealed in the crystals. The crystals that allowed the structure determination of the MaCCC domain at high resolution were obtained from protein that included the His_6-tag, which was not defined in the electron density. Although the asymmetric unit of this crystal form contains only one copy of the protein, a pair of interacting domains is related by two-fold crystallographic symmetry. A dimer of the protein as observed in this crystal form is shown in Figure 3. The two subunits bind via two discontinuous regions on the surface in a "head to tail" arrangement (Figure S6). This oligomeric structure brings

Structure
CCC Domain Structure

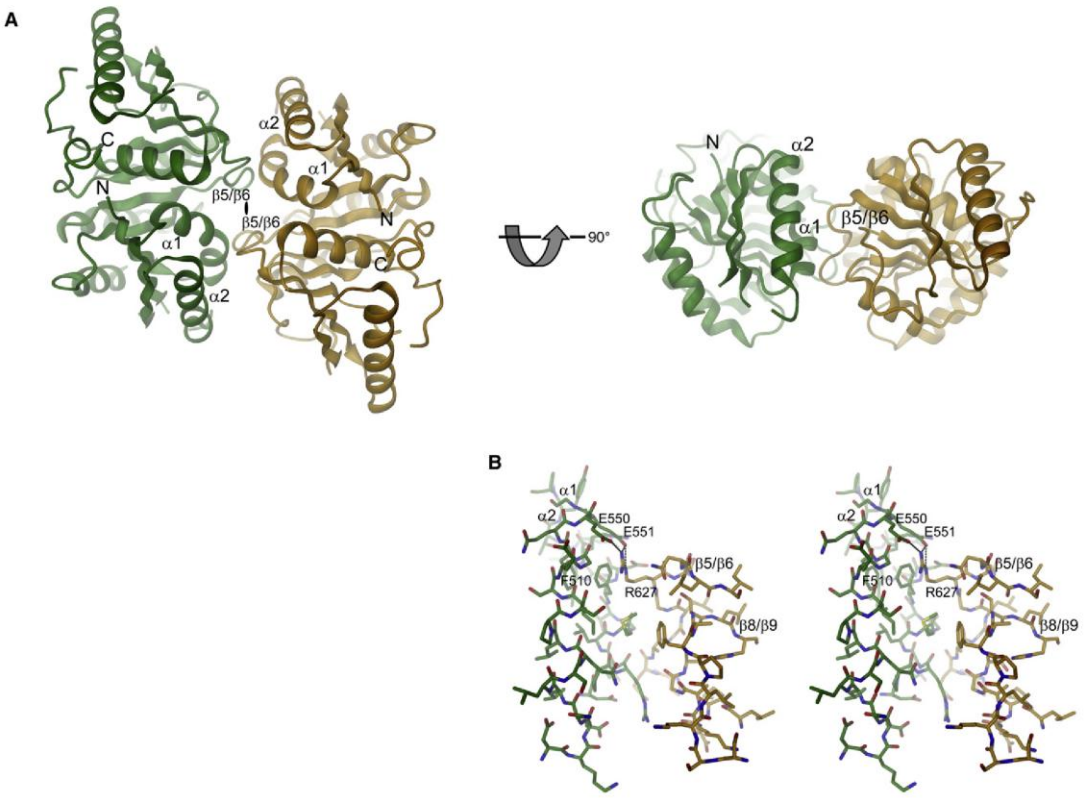


Figure 3. Oligomeric Organization of the MaCCC Domain

(A) Ribbon representation of the MaCCC domain dimer. The two subunits are colored in green and orange, respectively. Left: The view is from the outside of the cell; both termini and selected secondary structure elements contributing to the dimer interface are labeled. The position of the symmetry axis is indicated. Right: The view is parallel to the membrane.

(B) Stereo view of the interaction region. The residues of one of the two equivalent patches in the interaction interface are shown. The orientation and color coding is as in (A), right panel. The ionic interactions of Arg 627 are indicated as dashed lines. Selected residues are labeled.

residues located on α helices 1 and 2 in one subunit into proximity to residues located in the loop connecting β strands 5 and 6 in the second subunit. The dimeric protein forms a disk-like structure with overall dimensions of 65 × 55 × 35 Å that places the respective N and C termini of both subunits at the same side (Figures 3A and S6). The interactions bury 1579 $Å^2$ of the combined molecular surface, the equivalent of 6% of the total surface. Apart from main-chain contacts between glycine residues, most of the interactions are hydrophilic and involve ionizable residues as arginines and glutamates that form intersubunit salt bridges and hydrogen bonds (Figure 3B). Several features of the interface observed in the crystal form of the MaCCC domain are unusual for native protein-protein interactions. (1) The buried interface that includes only 6% of the combined molecular surface is comparably small. (2) The interactions are mainly hydrophilic and do not contain an extended hydrophobic component that is usually believed to be an important driving force for oligomerization. (3) Finally, the interacting residues are not conserved between different family members (Figures 2D and S2). It is thus not obvious whether this protein-protein interface represents the interaction between the domains in solution and in the context of the dimeric full-length transporter, or whether it merely shows contacts that were formed during crystallization of the MaCCC domain.

Dimeric Relationships in Different Crystal Forms

To probe whether the intersubunit interactions found in the $P3_221$ crystal form of the MaCCC domain would also be observed in different crystalline environments, we subjected the protein to broad crystallization screening. By using a construct of the protein where the His-tag was removed, we succeeded with structure determination of two novel crystal forms of space groups C222 and $I4_1$ with two and four copies of the protein in the asymmetric unit, respectively (Figure S7). The structure of the individual subunits is in all cases very similar to

Structure
CCC Domain Structure

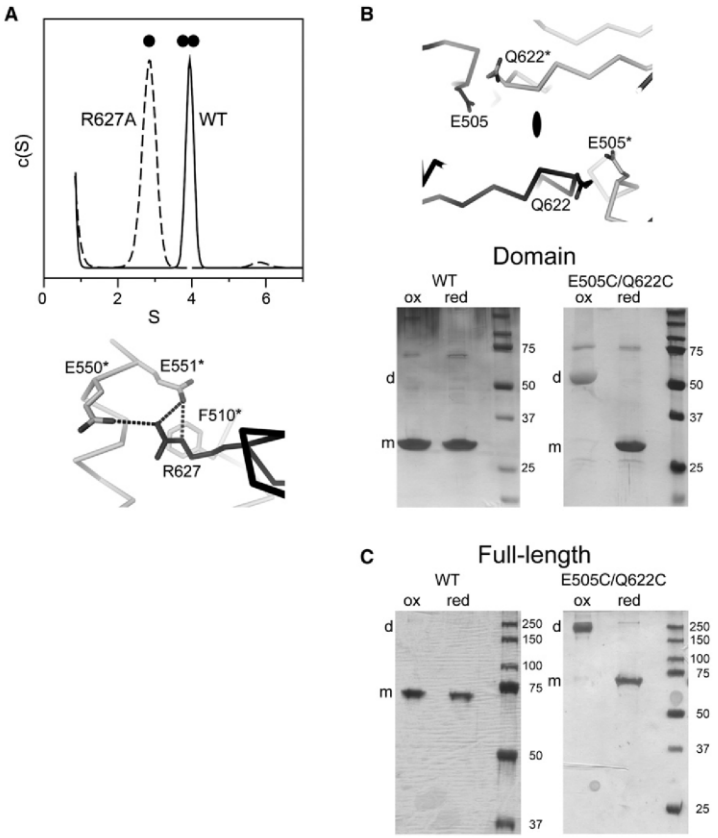


Figure 4. Analytical Ultracentrifugation and Crosslinking Experiments

(A) Distribution of the sedimentation coefficient (c[s]) as calculated from sedimentation velocity experiments for the WT MaCCC domain (—) and the domain mutant R627A (- - -). The His-tag was in both cases cleaved off. The oligomeric state (●, monomer; ●●, dimer) is indicated. A view of selected residues in the dimer interface shows the interactions of Arg 627 (*) with the neighboring subunit.

(B) Silver-stained SDS-PAGE of the purified MaCCC domain and the domain double mutant E505C/Q622C under oxidizing (ox) and reducing (red) conditions. The molecular weight of protein standards and the positions of the monomeric and dimeric proteins are indicated. A view of the dimer interface from the cytoplasm shows the relationship of the residues in the two subunits. Residues belonging to the second subunit are marked (*), and the dimer axis is indicated.

(C) Crosslinking of the purified full-length transporter. Silver-stained SDS-PAGE of full-length MaCCC (left) and the full-length double mutant E505C/Q622C (right). The molecular weight of protein standards and the positions of the monomeric and dimeric proteins are indicated.

the conformation observed in the original crystal form. Both new crystal forms contain pairs of tightly interacting domains that are related by two-fold symmetry with one arrangement corresponding to the original head-to-tail interaction that was found in the $P3_221$ crystal form. In the second arrangement, two subunits are interacting via one of the small faces of the trapezoid structure in a "head to head" interaction, thereby forming an elongated molecule that buries 881 Å^2 (i.e., 3.6%) of the combined molecular surface and that places the two termini at opposite faces of the oligomeric structure (Figure S8). Our attempts to investigate the interactions of pairs of MaCCC domains in different crystal forms has allowed us to locate the previously observed interface in a head-to-tail dimer, thus suggesting that this interaction is robust. The new crystal forms have also revealed a second head-to-head interaction interface that, although smaller than the initially identified arrangement, provides an alternative possibility for intersubunit interactions.

Probing the Interaction Interface by Mutagenesis

To investigate whether the intermolecular interactions observed in the different crystal forms of MaCCC would be preserved in solution, we mutated residues that are buried in the two different dimer interfaces. By inspection of the head-to-tail arrangement of the $P3_221$ crystal forms, we identified a prominent role for Arg 627 in intersubunit interactions (Figure 3B). The positively charged side chain that projects into the dimer interface forms a pair of salt bridges with two glutamate residues (Glu 550 and Glu 551) that are located at the C terminus of α helix 2 of the adjacent subunit, whereas its guanidinium group is involved in intersubunit stacking interactions with the aromatic side chain of a phenylalanine residue (Phe 510). The truncation of the side chain to alanine removes this network of interactions. The increase of the elution volume in size-exclusion chromatography of the mutant R627A already hinted at a smaller size of the protein. Subsequent analysis of the molecular weight by analytical ultracentrifugation revealed a slower sedimentation behavior that corresponds to a monomeric protein (Figure 4A). The experiments show that the truncation of Arg 627 leads to a disruption of the MaCCC domain dimer, and thus indicate that this residue plays an important role in intersubunit interactions in solution.

In a second set of experiments, we aimed to probe the proximity of residues in the dimer interface by cystein mutagenesis followed by the formation of an intermolecular disulfide bridge that crosslinks the two domains under mildly oxidizing conditions. Based on the structure, we chose a pair of residues that are at a suitable distance in the head-to-tail dimer to permit the formation of an intersubunit disulfide bond. The double mutant E505C/Q622C allowed us to express soluble protein that crosslinks two subunits under oxidizing conditions (Figure 4B). The mutation of either Leu 538 or Ile 559, which are both located at

the position of the two-fold symmetry axis in the head-to-head dimer, in contrast, failed to form crosslinks, and the double mutant L538A/I559A showed the same dimeric sedimentation behavior as the wild-type (WT) (Figure S8).

In both sets of experiments, the disruption of the head-to-tail dimer by the mutation of a critical residue in the interaction interface and the formation of site-specific disulfide bonds between pairs of close-by residues strongly suggest that the head-to-tail dimer, which was observed in several crystal forms, depicts the interaction between pairs of MaCCC domains in solution.

To investigate whether this interaction is also preserved in the context of the full-length protein, we expressed the transmembrane MaCCC transporter and its double mutant E505C/Q622C and studied the formation of the crosslink after extraction and purification (Figure 4C). The fact that the intersubunit crosslink was formed in the double mutant but not in the WT protein strongly suggests that the observed interface is preserved in the full-length protein where the cytoplasmic domains are attached to the C terminus of the transmembrane domain. Our results thus establish the structure of the cytoplasmic components of CCC transporters, support the previously observed homodimeric organization of the proteins, and reveal an interaction for the MaCCC domains that might be general for the family.

DISCUSSION

The structure of the cytoplasmic domain of MaCCC, a prokaryotic member of the cation-chloride cotransporter family, has shed first structural insight into this important family of secondary active ion transport proteins. The domain folds into a compact structure that consists of two structurally related subdomains that distantly resemble proteins of the USP family (Kvint et al., 2003). Although the sequences encoding for the C-terminal domains are only moderately conserved, there is evidence that the structure is representative of the entire family. The difference in size of this protein component, which ranges from about 300 residues in prokaryotes to up to 450 residues in eukaryotic transporters, is due to insertion of loop regions in the latter that likely do not fold into compact structures. The conservation of the structural organization of CCC transporter cytoplasmic domains is consistent with experiments on the human transporter NKCC2. In biochemical studies it was shown that two complementary fragments of the isolated C terminus, each containing the region of either one subdomain identified in the MaCCC transporter, can interact in solution and thus be recovered in pull-down experiments (Simard et al., 2004).

Different mammalian members of the CCC family have previously been identified to form homodimers and the C terminus was proposed to play an important role for dimer stability (Moore-Hoon and Turner, 2000; Parvin et al., 2007). In accordance with these studies, we found that the isolated domain of MaCCC forms stable dimers in solution. The structure of the protein in different crystal forms has allowed us to identify an oligomeric arrangement with an unusual protein interaction-interface that is comparably small, hydrophilic, and not conserved in sequence. Yet, different biochemical experiments on the isolated domains and the full-length transporter have shown strong evidence that the observed interface is preserved in solution and in the context of the full-length transporter. Because of the low sequence conservation in the buried interface, it is difficult to predict whether this mode of interaction is common to members of the CCC family. There are, however, structural constraints that limit the conformational freedom of these components in the full-length transporter. Each domain is connected to the C terminus of the transmembrane part via a short linker and probably interacts with residues in intracellular loops. It is thus remarkable that the region in the cytoplasmic domain that surrounds its N terminus harbors the most conserved patches on the surface of the molecule (Figure 2E). Because interactions with the transmembrane domain would restrict the accessible range of intersubunit interactions, it is conceivable that besides the structure of the C-terminal domain also its oligomeric organization is preserved throughout the family. In light of the poor conservation in the interface, it is noteworthy that previous studies have identified the cytoplasmic domains of CCC transporters to be critical elements in subunit recognition. For example, chimeras of the transmembrane part of NKCC1 and the domains of the closely related NKCC2 failed to dimerize with WT NKCC1, whereas they formed homodimers of chimeric subunits, thus suggesting that the domains not only significantly contribute to the dimer stability, but are also the main determinants for specificity and in that way prevent the formation of heterodimers between different members of the family (Parvin et al., 2007). The poor conservation of the interaction interface could thus reflect this demand on subunit specificity.

The role of the cytoplasmic domains for transporter function is currently not understood. In other families of transport proteins, however, such cytoplasmic components have been found to play an important role in the regulation of transport activity in response to the interaction with intracellular ligands (Meyer et al., 2007; Zagotta et al., 2003). Different experiments on members of the CCC transporters suggest that these folded protein components could play a similar role, but the mechanism of this process still awaits experimental clarification (Bergeron et al., 2006; Mercado et al., 2006; Pedersen et al., 2008). The results of our study provide a structural framework for the cytoplasmic domains of CCCs and can serve as basis for future investigations on a family that, despite its large physiological importance, still lacks detailed mechanistic insight.

EXPERIMENTAL PROCEDURES

Protein Preparation

The gene encoding the full-length cation-chloride cotransporter from *Methanosarcina acetivorans* (MaCCC, accession number MA4506) and a fragment encoding the cytoplasmic domain encompassing residues 474–758 were cloned into a pET28-b+ vector (Novagen) to generate constructs of the proteins followed by a C-terminal herpes simplex (HRV) 3C protease site (GE Healthcare) and a His_6-tag. A second construct of the MaCCC domain was similar but lacked the 3C protease site. For expression of the MaCCC domain *E. coli* BL21 (DE3) cells transformed with the expression construct were grown at 37°C in LB medium containing 50 μg/ml kanamycin to an OD_{600} of 0.6. Expression was induced by addition of 0.5 mM isopropyl-D-thiogalactopyranoside at 37°C for 3 hours. Cells were harvested and lysed with an Emulsiflex high-pressure homogenizer (Avestin) in 50 mM Tris-HCl (pH 8), 150 mM NaCl (buffer A) with addition of 1 mg/ml lysozyme, 20 μg/ml DNase, 1 μg/ml leupeptin, 1 μg/ml pepstatin, and 1 mM phenylmethyl sulphonyl fluoride. The lysate was cleared by centrifugation and the protein was purified by affinity-chromatography on a Ni-NTA column (Quiagen). For constructs

Structure
CCC Domain Structure

Table 1. Data Collection and Model Refinement Statistics

	Native	SeMet	C222	$P4_1$
Data Collection	Synchrotron	Synchrotron	Synchrotron	Synchrotron
Space group	$P3_221$	$P3_221$	C222	$P4_1$
Cell dimensions	68.0, 68.0, 137.0	68.3, 68.3, 137.3	46.3, 216.7, 140.8	105.4, 105.4, 257.4
$\alpha = \beta = \gamma$	90, 90, 120	90, 90, 120	90, 90, 90	90, 90, 90
Wavelength (Å)	0.964	0.979	1.0	1.0
Resolution (Å)	50-1.9	50-2.1	50-3.0	50-2.85
R_{sym}[a]	6.6 (38.3)	7.6 (26.9)	9.0 (26.7)	5.7 (33.8)
$I/\sigma I$	20.2 (3.6)	25.9 (4.9)	12.1 (3.0)	15.5 (2.5)
Completeness (%)	95.6 (74.6)	97.8 (80.8)	89.7 (77.0)	99.8 (99.2)
Mosaicity (°)	0.42	0.28	0.90	0.20
Refinement				
Resolution (Å)	10-1.9		10-3.0	10-2.85
No. of reflections	28 226		12 813	29 634
R_{work}[b] / R_{free}[c]	19.8 (23.2)		25.1 (30.9)	25.1 (30.3)
Number of protein atoms	2050		4100	8200
Number of waters	129			
Rmsd length (Å)	0.007		0.008	0.007
Rmsd angles (°)	1.06		1.25	1.04
Average B-factor	55.7 (54.6)		73.7	118

Values for the highest-resolution shell are given in parentheses.
[a] $Rsym = \Sigma|I_i - < I_i >| / \Sigma|I_i|$, where I_i is the scaled intensity of the i^{th} measurement and $< I_i >$ is the mean intensity for that reflection.
[b] $R_{work} = \Sigma||F_o|-|F_c|| / \Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.
[c] R_{free} was calculated using a randomly selected 10% sample of the reflection data omitted from refinement.

containing the 3C protease cleavage site, the protein was digested with HRV 3C protease (GE Healthcare) to remove the His_6-tag. After concentrating the protein, it was subjected to gel filtration on a Superdex 200 column (GE Healthcare). For preparation of selenomethionine-labeled protein, the bacterial culture was grown in minimal medium containing 50 mg/l selenomethionine.

For expression of full-length MaCCC, transformed *E. coli* BL21 (DE3) cells were grown in autoinduction media supplemented with 50 μg/ml kanamycin at 30°C overnight. Cells were harvested and lysed as described above. The lysate was cleared by low-spin centrifugation. Membranes were isolated by ultracentrifugation and the proteins were extracted in buffer A containing 1% n-decyl-β-D-maltoside (DM, Anatrace, Inc.). All subsequent purification steps were as described above but with addition of 5 mM DM in all buffer solutions.

Mutations were inserted with the QuikChange method (Stratagene) and confirmed by sequencing. Mutant proteins were purified following the same protocol as for WT. For crosslinking studies, all constructs had the His_6-tag removed. The respective disulfide bridges formed spontaneously upon exposure to air or after addition of 1mM $Na_2S_4O_6$ and were reduced by addition of β-mercaptoethanol. The proteins were analyzed by SDS-PAGE.

Analytical Ultracentrifugation

Sedimentation velocity experiments were performed with a Beckman model XL-I analytical ultracentrifuge with a 50-Ti rotor at 4°C. Double-sector centerpieces were filled with 400 μl protein samples and 420 μl buffer A. Data were acquired at 280 nm in continuous-scan mode in 0.003 cm intervals at a rotor speed of 40,000 rpm. Sedimentation analysis of the MaCCC cytoplasmic domain and the mutant R627A was performed at protein concentrations of 15 μM. The His_6-tag was removed unless stated explicitly. Data analysis was performed with the c(S) module of Sedfit (Schuck, 2000; Schuck et al., 2002). The buffer parameters, partial-specific volume of the protein, and the corrected sedimentation coefficients $S°_{20w}$ were calculated by using Sednterp (Laue et al., 1992). For data used in this analysis, the root-mean-square deviations of all the fits were between 0.005 and 0.01.

Crystallization and Crystal Preparation

All crystals were obtained by sitting drop vapor diffusion. Purified protein with a concentration of 8–15 mg/ml was mixed in a 1:1 ratio with reservoir solution and was equilibrated against the reservoir. Crystals of the MaCCC domain in a trigonal crystal form were grown from protein construct lacking the 3C protease site and thus had the His_6-tag attached. The reservoir solution contained 100 mM MES (pH 6.5) and 1.1 M MgSO4. Crystals grew after 1–3 days at 4°C in space group $P3_221$ with one copy of the MaCCC domain in the asymmetric unit. Crystals of the MaCCC domain in the orthorhombic and tetragonal crystal forms were grown from protein where the His_6-tag was removed. Crystals in the orthorhombic crystal form were grown in a reservoir solution containing 100 mM sodium acetate (pH 4.6) and 750 mM $MgCl_2$. Crystals grew after 6 days at 20°C in space group C222 with two molecules in the asymmetric unit. Crystals of the tetragonal crystal form were grown from reservoir solution containing 100 mM Tris (pH 8.5) and 3.9 M ammonium chloride. Crystals grew after 4 days at 20°C in space group $I4_1$ with four copies in the asymmetric unit. For cryoprotection, the crystals in all three crystal forms were successively transferred into mother liquor containing increasing amounts of glycerol to a final concentration of 35%, flash-frozen in liquid propane, and stored in liquid nitrogen.

Crystallography

Data sets were collected on frozen crystals on the X06SA beamline at the Swiss Light Source of the Paul Scherrer Institut on a Mar225 CCD detector (Marresearch). The data were indexed, integrated, and scaled with the programs XDS (Kabsch, 1993), DENZO, and SCALEPACK (Otwinowski and Minor, 1997), and further processed with CCP4 programs (Table 1) (CCP4, 1994). The structure of the $P3_221$ crystal form was determined by the SAD method using the anomalous contributions of Se-atoms in a crystal containing protein where methionine residues were replaced by selenomethionine. The Se-sites were identified with SHELX C and D (Pape and Schneider, 2004; Schneider and Sheldrick, 2002) and refined with SHARP (de La Fortelle and Bricogne, 1997). Phases were improved by solvent flattening in DM (Cowtan, 1994). The model was built in O and refined by simulated annealing with the program CNS (Brunger et al., 1998) alternated with inspection and manual

rebuilding in O. In later stages, the refinement proceeded in a native dataset at 1.9 Å resolution with PHENIX (Adams et al., 2002). Individual B-factors were refined and water molecules were placed into residual electron density. R_{free} was calculated by selecting 10% of the reflection data that were omitted in refinement. Twenty-eight residues at the N terminus and in loop regions were not defined in the electron density and are thus not included in the model. The final has R/R_{free} values of 19.8% and 23.2%, good geometry and no outliers in the Ramachandran plot. The structures of the C222 and $I4_1$ crystal forms were determined by molecular replacement with PHASER (McCoy et al., 2007) using the refined MaCCC domain structure as search model and refined in PHENIX maintaining tight NCS constraints.

ACCESSION NUMBERS

Coordinates have been deposited with the Protein Data Bank under code 3G40.

SUPPLEMENTAL DATA

Supplemental Data include eight figures and Supplemental References and can be found with this article online at http://www.cell.com/structure/supplemental/S0969-2126(09)00090-2.

ACKNOWLEDGMENTS

We would like to thank B. Blattmann and A. Haisch for assistance with crystal screening, C. Schulze-Briese and the staff of the X06SA beamline for support during data collection, the protein analysis group at the functional genomics center of the University of Zurich for help with mass spectrometry, and members of the Dutzler lab for help in all stages of the project. Data collection was performed at the Swiss Light Source of the Paul Scherrer Institute. The research leading to these results has received funding from a grant from the National Center for Competence in Research in Structural Biology. S.W. is affiliated with the Molecular Life Sciences Ph.D. program of the University/ETH Zurich. The authors declare that they have no competing financial interest.

Received: December 16, 2008
Revised: February 4, 2009
Accepted: February 5, 2009
Published: April 14, 2009

REFERENCES

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. *58*, 1948–1954.

Barabote, R.D., Tamang, D.G., Abeywardena, S.N., Fallah, N.S., Fu, J.Y., Lio, J.K., Mirhosseini, P., Pezeshk, R., Podell, S., Salampessy, M.L., et al. (2006). Extra domains in secondary transport carriers and channel proteins. Biochim. Biophys. Acta *1758*, 1557–1579.

Bergeron, M.J., Gagnon, E., Caron, L., and Isenring, P. (2006). Identification of key functional domains in the C terminus of the K+-Cl- cotransporters. J. Biol. Chem. *281*, 15959–15969.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

CCP4. (1994). Collaborative Computational Project Nr. 4. The CCP4 Suite: Programs for X-ray crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Cowtan, K. (1994). An automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography *31*, 34–38.

de La Fortelle, E., and Bricogne, G. (1997). Methods in enzymology. In Methods in Enzymology, C.W. Carter and R.M. Sweet, eds. (New York: Academic Press), pp. 492–494.

Dowd, B.F., and Forbush, B. (2003). PASK (proline-alanine-rich STE20-related kinase), a regulatory kinase of the Na-K-Cl cotransporter (NKCC1). J. Biol. Chem. *278*, 27347–27353.

Gamba, G. (2005). Molecular physiology and pathophysiology of electroneutral cation-chloride cotransporters. Physiol. Rev. *85*, 423–493.

Gimenez, I., and Forbush, B. (2005). Regulatory phosphorylation sites in the NH2 terminus of the renal Na-K-Cl cotransporter (NKCC2). Am. J. Physiol. Renal Physiol. *289*, F1341–F1345.

Hebert, S.C., Mount, D.B., and Gamba, G. (2004). Molecular physiology of cation-coupled Cl- cotransport: the SLC12 family. Pflugers Arch. *447*, 580–593.

Isenring, P., and Forbush, B. (2001). Ion transport and ligand binding by the Na-K-Cl cotransporter, structure-function studies. Comp. Biochem. Physiol. A Mol. Integr. Physiol. *130*, 487–497.

Isenring, P., Jacoby, S.C., Chang, J., and Forbush, B. (1998a). Mutagenic mapping of the Na-K-Cl cotransporter for domains involved in ion transport and bumetanide binding. J. Gen. Physiol. *112*, 549–558.

Isenring, P., Jacoby, S.C., and Forbush, B., 3rd. (1998b). The role of transmembrane domain 2 in cation transport by the Na-K-Cl cotransporter. Proc. Natl. Acad. Sci. USA *95*, 7179–7184.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown.symmetry and cell constants. J. Appl. Crystallogr. *26*, 795–800.

Kvint, K., Nachin, L., Diez, A., and Nystrom, T. (2003). The bacterial universal stress protein: function and regulation. Curr. Opin. Microbiol. *6*, 140–145.

Laue, T., Shah, B., Ridgeway, T., and Pelletier, S. (1992). Computer aided interpretation of analytical sedimentation data for proteins, S. Harding, A. Rowe, and J. Horton, eds. (Cambridge, UK: The Royal Society of Chemistry).

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Mercado, A., Broumand, V., Zandi-Nejad, K., Enck, A.H., and Mount, D.B. (2006). A C-terminal domain in KCC2 confers constitutive K+-Cl- cotransport. J. Biol. Chem. *281*, 1016–1026.

Meyer, S., Savaresi, S., Forster, I.C., and Dutzler, R. (2007). Nucleotide recognition by the cytoplasmic domain of the human chloride transporter ClC-5. Nat. Struct. Mol. Biol. *14*, 60–67.

Moore-Hoon, M.L., and Turner, R.J. (2000). The structural unit of the secretory Na+-K+-2Cl- cotransporter (NKCC1) is a homodimer. Biochemistry *39*, 3718–3724.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *267*, 307–326.

Pape, T., and Schneider, T.R. (2004). HKL2MAP: a graphical user interface for phasing with SHELX programs. J. Appl. Crystallogr. *37*, 843–844.

Park, J.H., and Saier, M.H., Jr. (1996). Phylogenetic, structural and functional characteristics of the Na-K-Cl cotransporter family. J. Membr. Biol. *149*, 161–168.

Parvin, M.N., Gerelsaikhan, T., and Turner, R.J. (2007). Regions in the cytosolic C-terminus of the secretory Na(+)-K(+)-2Cl(-) cotransporter NKCC1 are required for its homodimerization. Biochemistry *46*, 9630–9637.

Payne, J.A., Rivera, C., Voipio, J., and Kaila, K. (2003). Cation-chloride co-transporters in neuronal communication, development and trauma. Trends Neurosci. *26*, 199–206.

Pedersen, M., Carmosino, M., and Forbush, B. (2008). Intramolecular and intermolecular fluorescence resonance energy transfer in fluorescent protein-tagged Na-K-Cl cotransporter (NKCC1): sensitivity to regulatory conformational change and cell volume. J. Biol. Chem. *283*, 2663–2674.

Schneider, T.R., and Sheldrick, G.M. (2002). Substructure solution with SHELXD. Acta Crystallogr. D Biol. Crystallogr. *58*, 1772–1779.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys. J. *78*, 1606–1619.

Schuck, P., Perugini, M.A., Gonzales, N.R., Howlett, G.J., and Schubert, D. (2002). Size-Distribution Analysis of Proteins by Analytical Ultracentrifugation: Strategies and Application to Model Systems. Biophys. J. *82*, 1096–1111.

Simard, C.F., Brunet, G.M., Daigle, N.D., Montminy, V., Caron, L., and Isenring, P. (2004). Self-interacting domains in the C terminus of a cation-Cl- cotransporter described for the first time. J. Biol. Chem. *279*, 40769–40777.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

Zagotta, W.N., Olivier, N.B., Black, K.D., Young, E.C., Olson, R., and Gouaux, E. (2003). Structural basis for modulation and agonist specificity of HCN pacemaker channels. Nature *425*, 200–205.

2.4.1.2.1 Supplementary data

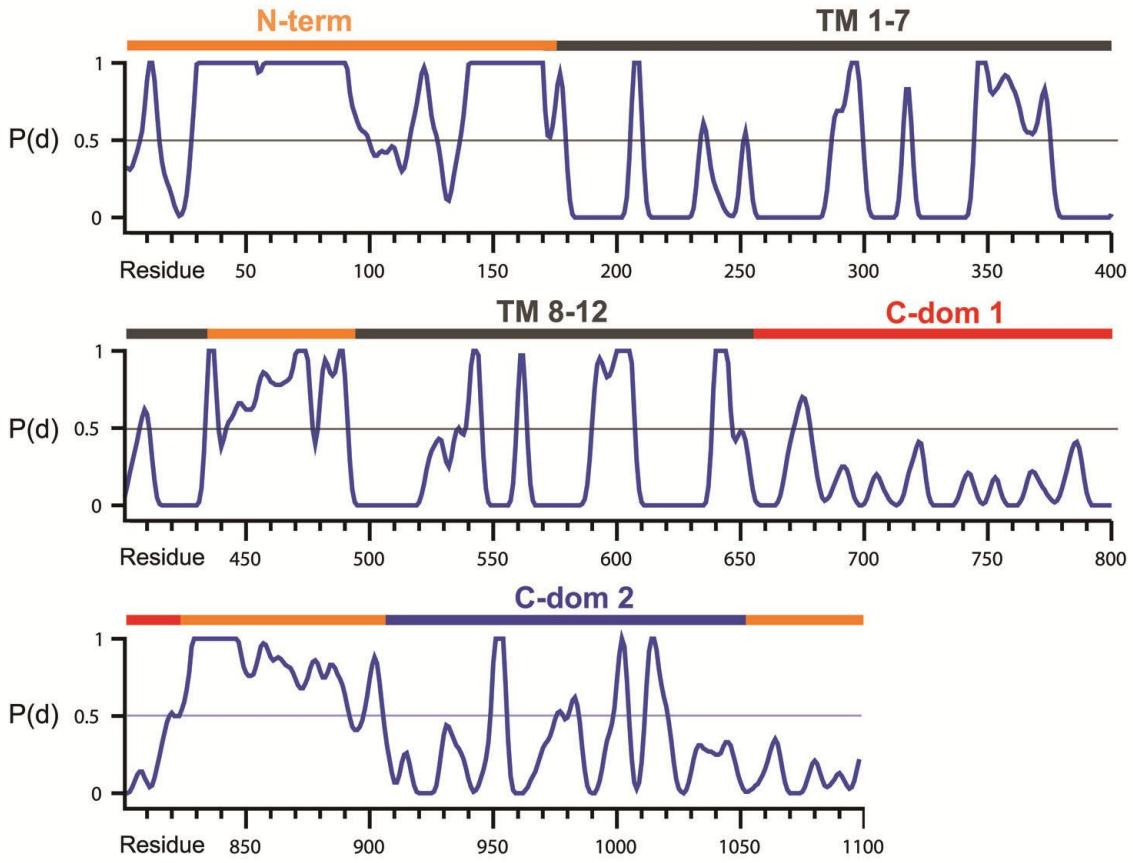


Fig S-1 Unstructured regions in human NKCC2.
The sequence of human NKCC2 (accession number SLC12A1) was analyzed with respect to the probability to fold into a defined conformation. P(d) was calculated with the program DRIPPRED (MacCallum 2004) and describes the propensity for disorder. Residues with values above P(d) = 0.5 are likely to be unstructured. Different regions are indicated above and shown in unique colors (transmembrane regions in grey, insertions with respect to MaCCC in orange and the two halves of the cytoplasmic domain in red and green respectively). Insertions at the N-terminus and in loop regions show a large propensity for being unstructured.

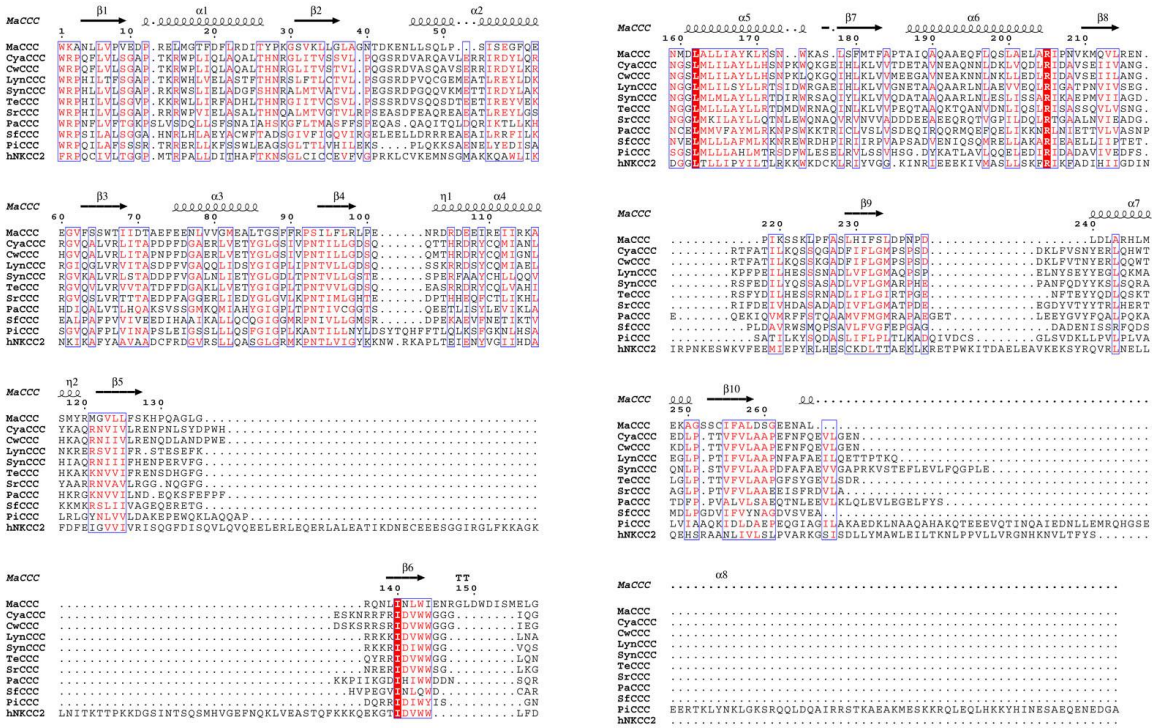


Fig S-2 Sequence conservation in the C-terminal domain
Multiple sequence alignment of the C-terminus of different prokaryotic CCC transporters and human NKCC2. Secondary structure and numbering of the MaCCC domain are shown above. Regions of high homology are colored red. Amino-acid sequences are: MaCCC; *Methanosarcina acetivorans C2A* (MA4506), CyaCCC; *Cyanothece sp. ATCC 51142*, (cce_0839), CwCCC; *Crocosphaera watsonii WH 8501* (EAM51764), LynCCC; *Lyngbya sp. PCC 8106* (EAM34907), SynCCC; *Synechococcus sp. PCC 7002* (SYNPCC7002_A1963), TeCCC; *Trichodesmium erythraeum IMS101* (Tery_1903), SrCCC; *Salinibacter ruber DSM 13855* (SRU_2259), PaCCC; *Candidatus Protochlamydia amoebophila UWE25* (GenBank pc0822), SfCCC; *Syntrophobacter fumaroxidans MPOB* (Sfum_3620), PiCCC, *Psychromonas ingrahamii 37* (Ping_1733), hNKCC2; *homo sapiens* (SLC12A1). The Alignment was made with ClustalW (Thompson, Higgins et al. 1994).

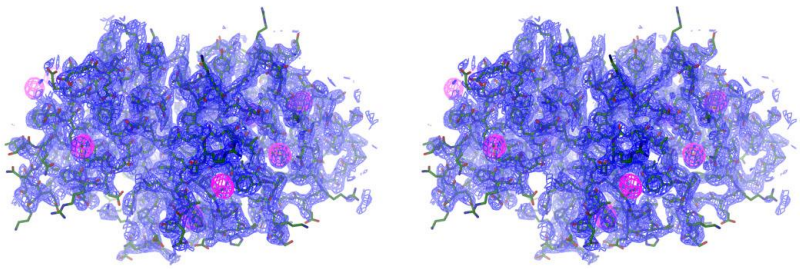


Fig S-3 Experimental electron density
Stereo view of electron density (calculated at 2.5 Å resolution and contoured at 1 σ, blue) shown superimposed on the refined structure of the MaCCC domain. The map was calculated from native amplitudes and solvent flattened SAD phases. Anomalous difference electron density of Se atoms (calculated at 3 Å resolution and contoured at 5 σ) is shown in magenta.

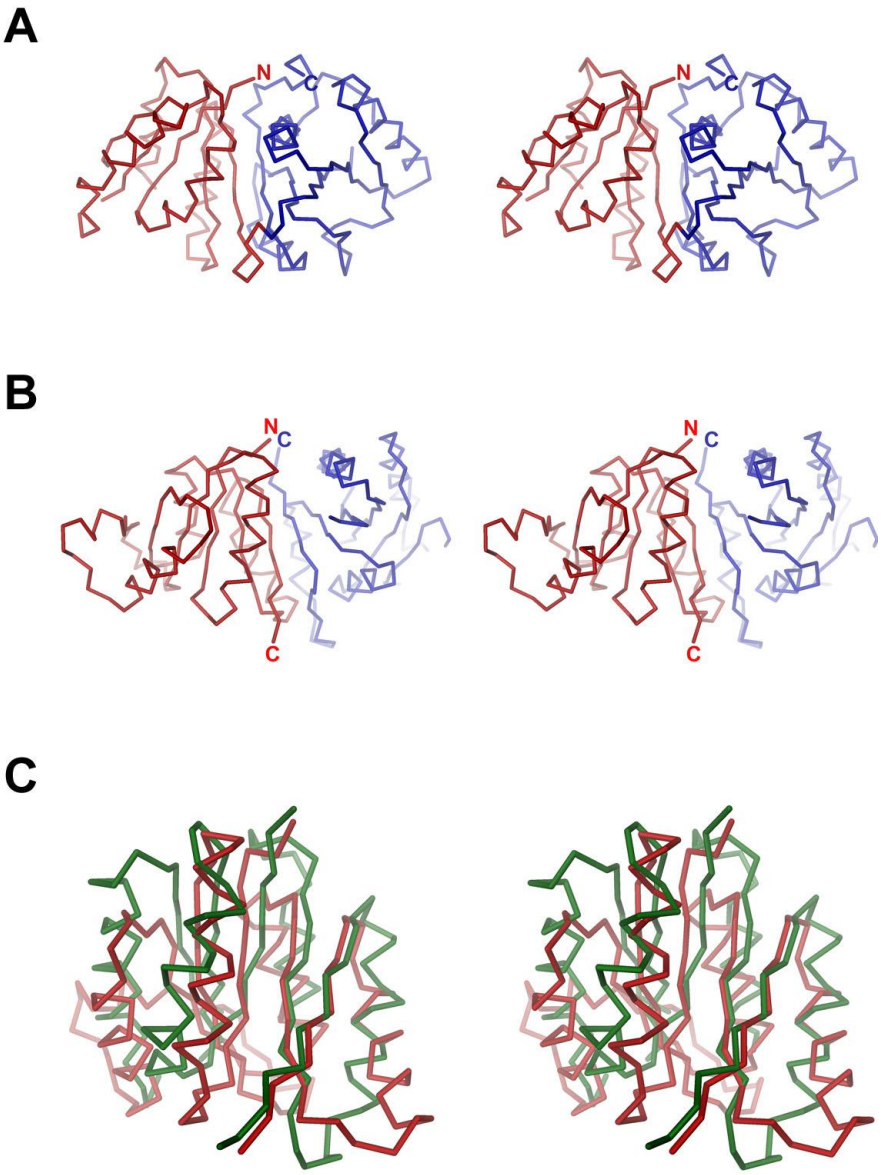


Fig S-4 Structural homology to USP domain proteins
(A) Stereo view of a Cα-trace of the MaCCC domain subunit. The two halves of the protein are colored in red and blue respectively. The N- and C-terminus is indicated. **(B)** Cα-trace of an interacting pair of USP domains (PDB code 2GM3). The two chains are colored in red and blue respectively, N- and C-termini are labeled. **(C)** Superposition of the Ca traces of the N-terminal sub-domain of the MaCCC C-terminus (green) and a single USP domain (red).

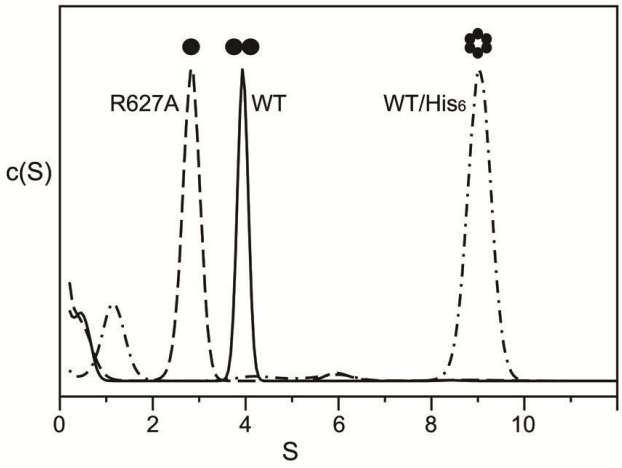


Fig S-5 Analytical ultracentrifugation of the MaCCC domain
Distribution of the sedimentation coefficient (c[s]) as calculated from sedimentation velocity experiments. The graph shows the distribution for the WT MaCCC domain (—) and the domain mutant R627A (- - - -) with cleaved-off His-tag, and WT still containing the His-tag (- · - · - · -). The oligomeric states are indicated.

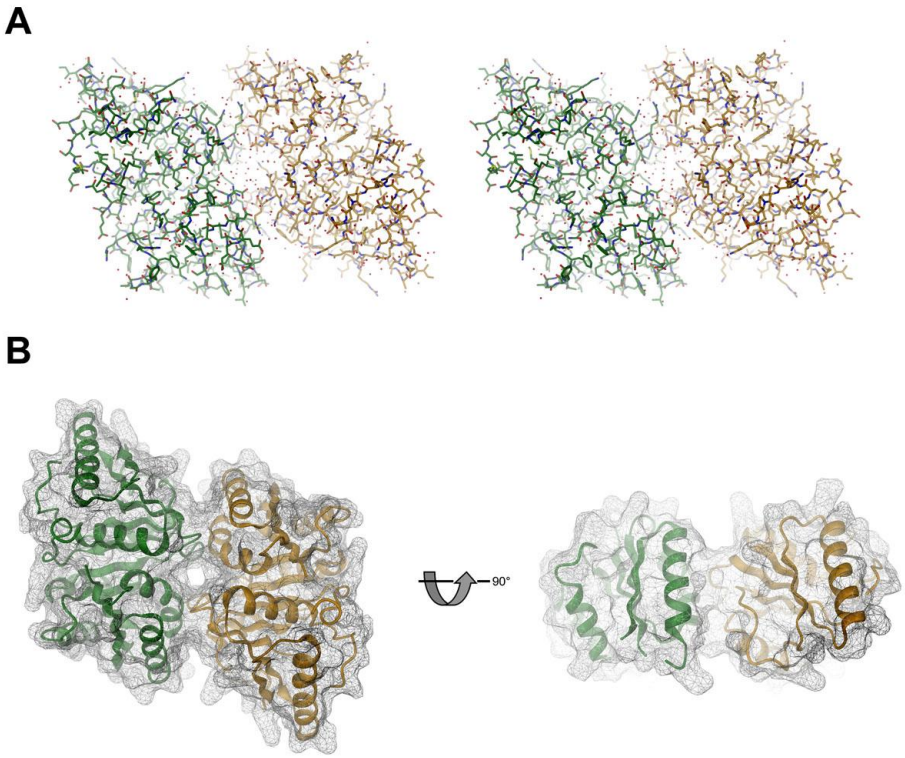


Fig S-6 MaCCC domain 'head to tail' dimer
(A) Stereo view of the MaCCC domain dimer. The view is along the two-fold axis from the extracellular side. The two chains are colored in green and orange respectively, ordered water molecules are shown as red spheres. (B) View of the dimer in two orientations. The solvent accessible surface (calculated with MSMS (Sanner, Olson et al. 1996)) is shown superimposed on a ribbon representation of the protein chain. The relationship between the orientations is indicated

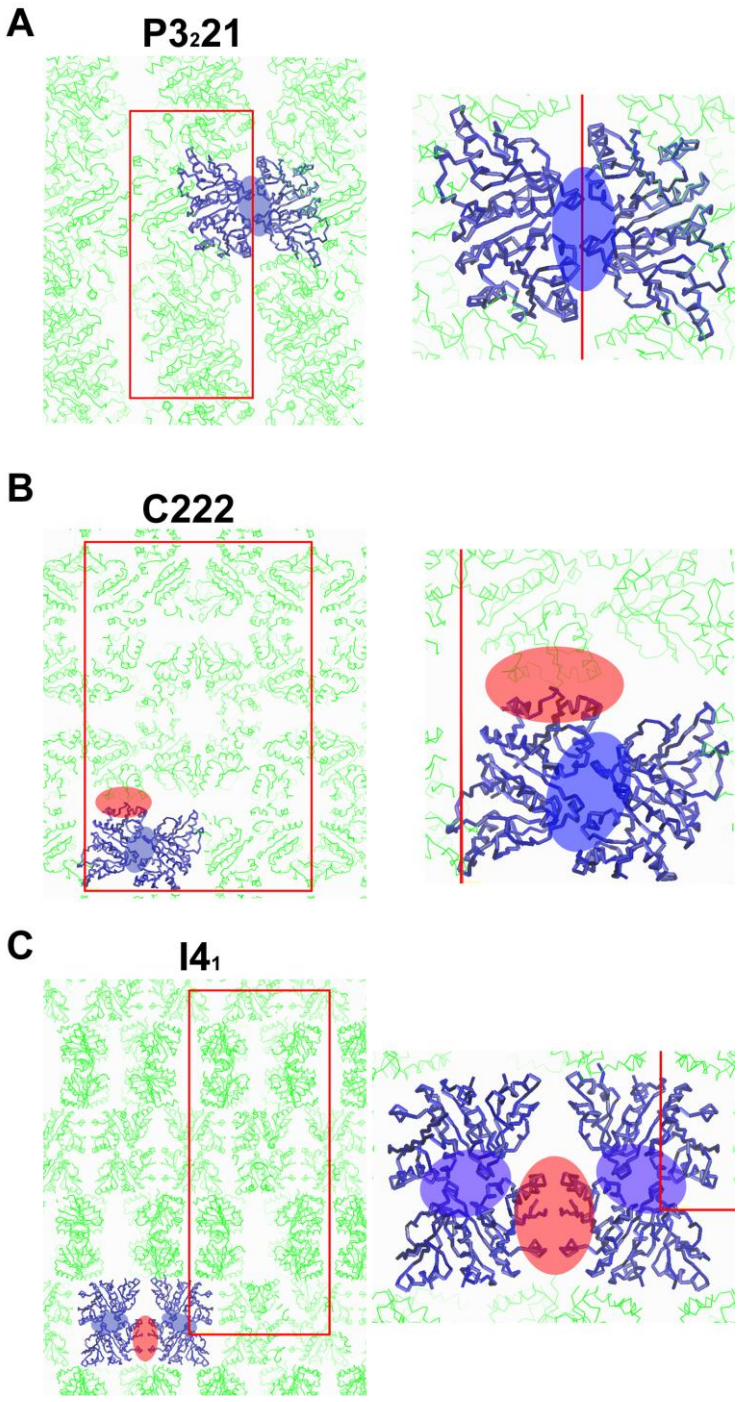


Fig S-7 Crystal packing
Two-fold relationships of tightly interacting pairs of MaCCC domains in different crystal forms. The interaction interfaces are highlighted ('head to tail' interaction in blue, 'head to head' interaction in red).The right panels show a magnified section of the packing. The unit cells are drawn in red. **(A)** $P3_221$ crystal form. A pair of proteins forming a 'head to tail dimer' related by crystallographic symmetry is shown in blue. **(B)** C222 crystal form. The asymmetric unit containing a pair of MaCCC domains related in the 'head to tail' dimer is shown in blue. The interface in a 'head to head' dimer formed by crystallographic symmetry is indicated in red. **(C)** $I4_1$ crystal form. The asymmetric unit consisting of four copies of the protein is shown in blue. Two fold relationships within the asymmetric unit for 'head to tail' and 'head to head' dimers are indicated in blue and red respectively.

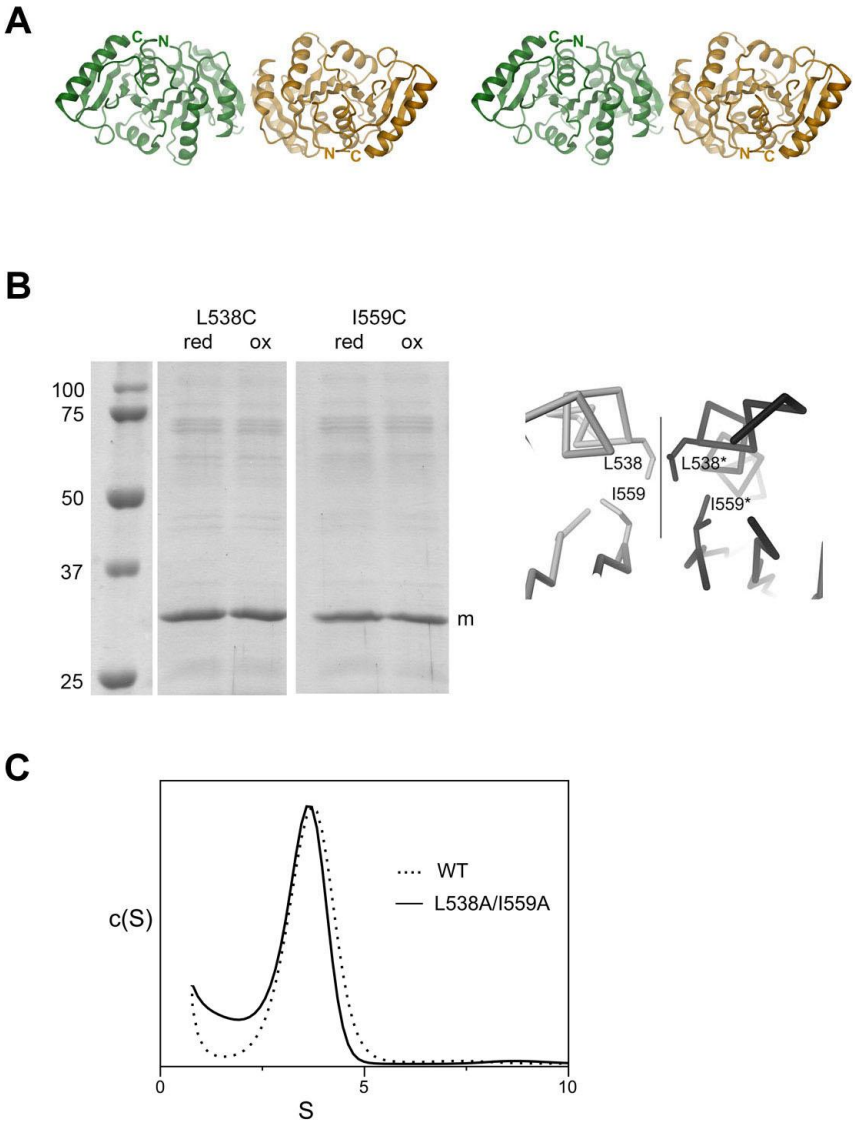


Fig S-8 Structure of the 'head to head' dimer
(A) Stereo view of a ribbon representation of a 'head to head' dimer. The subunits are colored in green and orange respectively, N- and C-termini are labeled. The view is along the two-fold axis. **(B)** Crosslinking attempts of adjacent residues in the ‘head to head’ dimer interface. Coomassie stained SDS-PAGE of the purified MaCCC domain mutants L538C and I559C under oxidizing (ox) and reducing (red) conditions. The molecular weight of protein standards and the position of the monomeric protein are indicated. A view of the dimer interface perpendicular to the two-fold axis shows the relationship of mutated amino acids. Residues belonging to the second subunit are marked (*), the two-fold axis is indicated. **(C)** Distribution of the sedimentation coefficient (c[s]) as calculated from sedimentation velocity experiments. The graph shows the distribution for the WT MaCCC domain (·····) and the double mutant L538A/I559A with cleaved-off His-tag. Both proteins sediment as dimers.

2.4.2 Materials and Methods

Protein Preparation. The gene encoding the cytoplasmic domain cation-chloride cotransporter from *Methanosarcina acetivorans* (MaCCC, accession number MA4506) encompassing residues 474-758 was cloned into a pET28-b+ vector (Novagen) to generate constructs of the proteins followed by a C-terminal *Herpes simplex* (HRV) 3C protease site (GE Healthcare) and a His_6-tag. A second construct of the MaCCC domain was similar but lacked the 3C protease site. For expression of the MaCCC domain *E. coli* BL21 (DE3) cells transformed with the expression construct were grown at 37 °C in LB medium containing 50 µg/ml kanamycin to an OD_{600nm} of 0.6. Expression was induced by addition of 0.5 mM isopropyl-D-thiogalactopyranoside (IPTG) at 37 ºC for three hours. Cells were harvested and lysed with an Emulsiflex high pressure homogenizer (Avestin) in 50 mM Tris-HCl (pH 8), 150 mM NaCl (buffer A) with addition of 1 mg/ml lysozyme, 20 µg/ml DNase, 1 µg/ml leupeptin, 1 µg/ml pepstatin and 1 mM phenylmethyl sulphonyl fluoride (PMSF). The lysate was cleared by centrifugation and the protein was purified by affinity-chromatography on a Ni-NTA column (Quiagen). For constructs containing the 3C protease cleavage site, the protein was digested with HRV 3C protease (GE Healthcare) to remove the His_6-tag. After concentrating, the protein was subjected to gel-filtration on a Superdex 200 column (GE Healthcare). For preparation of seleno-methionine labeled protein, the bacterial culture was grown in minimal medium containing 50 mg/l seleno-methionine.

Mutation and Disulfide Bridge Formation. Mutations were inserted with the QuikChange method (Stratagene) and confirmed by sequencing. Mutant proteins were purified following the same protocol as for WT. For crosslinking studies all constructs had the His_6-tag removed. The respective disulfide bridges formed spontaneously upon exposure to air or after addition of 1mM $Na_2S_4O_6$ and were reduced by addition of β-mercaptoethanol. The proteins were analyzed by SDS-PAGE.

Analytical Ultracentrifugation. Sedimentation velocity experiments were performed with a Beckman model XL-I analytical ultracentrifuge with a 50-Ti rotor at 4°C. Double-sector centerpieces were filled with 400 µl of protein samples and 420 µl of buffer A. Data were acquired at 280 nm in continuous-scan mode in 0.003 cm intervals at a rotor speed of 40,000 rpm. Sedimentation analysis of the MaCCC cytoplasmic domain and the mutant R627A was performed at protein concentrations of 15 µM. The His_6-tag was removed unless stated explicitly. Data analysis was performed with the c(S) module of Sedfit (Schuck 2000; Schuck, Perugini et al. 2002). The buffer parameters, partial-specific volume of the protein, and the corrected sedimentation coefficients $S°_{20W}$ were calculated by using Sednterp. For data used in this analysis, the root-mean-square deviations of all the fits were between 0.005 and 0.01.

Crystallization and Crystal Preparation. All crystals were obtained by sitting drop vapor diffusion. Purified protein with a concentration of 8-15 mg/ml was mixed in a 1:1 ratio with reservoir solution and was equilibrated against the reservoir. Crystals of the MaCCC domain in a trigonal crystal form were grown from protein construct lacking the 3C protease site and thus had the His_6-tag attached. The reservoir solution contained 100 mM MES (pH6.5), 1.1 M MgSO4. Crystals grew after 1-3 days at 4 °C in space group $P3_221$ with one copy of the MaCCC domain in the asymmetric unit. Crystals of the MaCCC domain in the orthorhombic and tetragonal crystal forms were grown from protein where the His_6-tag was removed. Crystals in the orthorhombic crystal form were grown in a reservoir solution containing 100 mM sodium acetate (pH4.6) and 750 mM $MgCl_2$. Crystals grew after six days at 20 °C in space group C222 with two molecules in the asymmetric unit. Crystals of the tetragonal crystal form were grown from reservoir solution containing 100 mM Tris (pH 8.5) and 3.9 M ammonium chloride. Crystals grew after four days at 20 °C in space group $I4_1$ with four copies in the asymmetric unit. For cryoprotection, the crystals in all three crystal forms were successively transferred into mother liquor containing increasing amounts of glycerol to a final concentration of 35%, flash-frozen in liquid propane and stored in liquid nitrogen.

Crystallography. Data sets were collected on frozen crystals on the X06SA beamline at the Swiss Light Source (SLS) of the Paul Scherrer Institut (PSI) on a Mar225 CCD detector (Marresearch). The data were indexed, integrated and scaled with the programs XDS (Kabsch 1993), DENZO and SCALEPACK (Otwinowski and Minor 1997) and further processed with CCP4 programs (Table 1) (CCP4 1994). The structure of the $P3_221$ crystal form was determined by the SAD method using the anomalous contributions of Se-atoms in a crystal containing protein where methionine residues were replaced by selenomethionine. The Se-sites were identified with SHELX C and D (Schneider and Sheldrick 2002; Pape and Schneider 2004) and refined with SHARP (de La Fortelle and Bricogne 1997). Phases were improved by solvent flattening in DM (Cowtan 1994). The model was built in O and refined by simulated annealing with the program CNS (Brunger, Adams et al. 1998) alternated with inspection and manual rebuilding in O. In later stages the refinement proceeded in a native dataset at 1.9 Å resolution with PHENIX (Adams, Grosse-Kunstleve et al. 2002). Individual B-factors were refined and water molecules were placed into residual electron density. R_{free} was calculated by selecting 10% of the reflection data that were omitted in refinement. 28 residues at the N-terminus and in loop regions were not defined in the electron density and are thus not included in the model. The final has R/R_{free} values of 19.8 % and 23.2 %, good geometry and no outliers in the Ramachandran plot. The structures of the C222 and $I4_1$ crystal forms were determined by molecular replacement with PHASER (A. J. McCoy 2007) using the refined MaCCC domain structure as search model and refined in PHENIX maintaining tight NCS constraints.

3 General Discussion and Outlook

In this work the structural organization of CCC transporters was investigated by biochemical methods and X-ray crystallography. Several prokaryotic homologues were identified and selected for expression, purification and crystallization studies. For all transporters three different expression constructs were designed encompassing either the full-length protein (section 2.3), the C-terminal domain (section 2.4) or the membrane domain (4.1). After broad expression screening and biochemical characterization the homologue of the archaea *Methanosarcina acetivorans*, MaCCC, was identified as the only protein where the three protein constructs could be expressed and purified in a stable state and in sufficient amounts for structural investigations. The The C-terminal domain of MaCCC allowed the structure determination of this putative regulatory subunit and the analysis of its oligomeric state. From different crystal forms two potential dimer interfaces were obtained, one of which was identified with biochemical tools to correspond to the native interface. The dimeric organization observed in the domain was shown to be conserved in the context of the full-length protein by MALS and SDS-PAGE. Although promising progress towards the structural and functional characterization of prokaryotic CCCs was reached, the final goal of the structure determination of a full-length CCC transporter remains very challenging.

Besides the improvement of expression, purification and crystallization, a future focus will be on functional investigations and the establishment of transport assays. MaCCC is clearly a very attractive protein for such studies since it is currently the only family member that has been purified in large quantities. The fact that the transport is electroneutral presents a challenge since it excludes most electrophysiological methods and it makes the transport independent of the membrane potential. However, insights into the transport properties and substrate specificity can potentially guide future purification and crystallization experiments.

To overcome the inherent instability of the family in solution stabilizing compounds have to be explored further. These binders can be small molecules as well as high affinity proteinaceous binders and should stabilize one conformation of the transporter. Specific binders are also relevant with respect to potential ligands binding to the cytoplasmic domain and regulating transport that are currently unknown. Due to the lack of structural information on the full-length protein the binders have to be selected from libraries. Potential binding molecules to the cytoplasmic domains could also be identified by virtual docking of possible substrates. The binding could either be detected by fluorescence methods, or isothermal titration calorimetry (ITC), or the modulation of transport observed in a transport assay. ITC would be particularly suited for screening

with the isolated C-terminal domain. The generation of suitable binding proteins to the transporter will be of large importance for future crystallization experiments. First steps in this direction have been described by the selection of specific antibodies in mouse. Since these experiments, however, did not allow the identification of epitope specific antibodies, which are believed to be a prerequisite for such co-crystallization experiments other approaches will have to be investigated. Due to the instability of the proteins future attempts for the generation of binders should be done *in vitro* by *e.g.* display methods. This offers the chance to control the state of the protein and rapidly produce numerous binders for defined conformations oligomeric states and construct size.

The work describe in this thesis has provided first structural information on the CCC family. A model of the current state of knowledge is summarized in Fig 3-1. The structure of the cytoplasmic domain was determined and has revealed a novel structure for a regulatory unit that might potentially interact with ligands to regulate transport in response to cellular signals. The identification of the ligand would provide novel important insight into the physiological processes that CCC transporters are involved in. The dimeric organization and the interface were confirmed with biochemical methods. The orientation of the C-terminal domain-dimer towards the membrane domain is based on the amino acid conservation mapping on the surface of the cytosolic domain. The identification of MaCCC as a well expressed and stable prokaryotic CCC transporter has established an important model system for future structural and functional

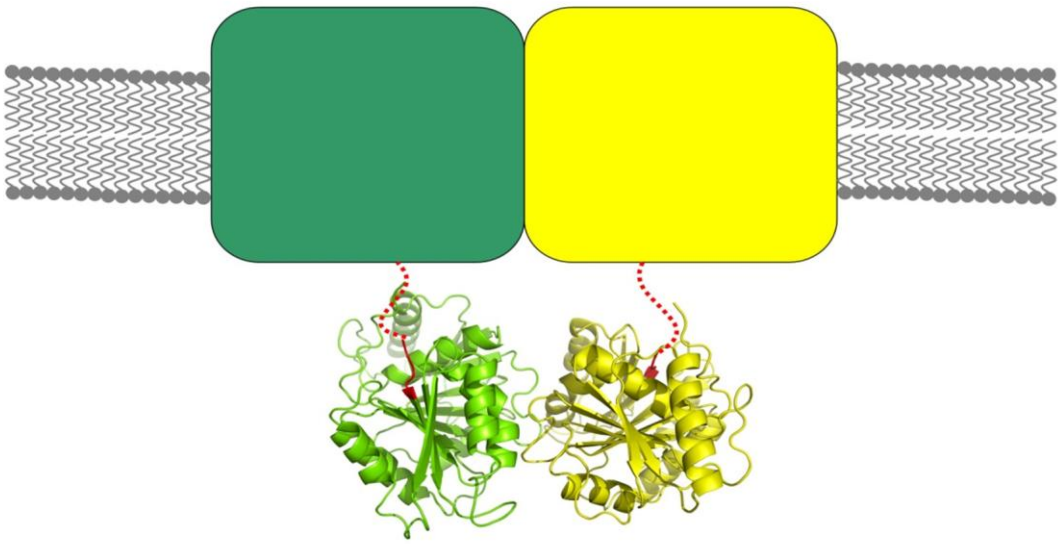


Fig 3-1 Model for CCC Organization
The C-terminal domains of MaCCC (green and yellow structures in cartoon) are connected to the membrane part (green and yellow box) with a linker (red). The structure of the membrane part (green and yellow boxes) are not determined yet. The membrane is depicted in grey.

investigation on a protein family that despite its importantance is still poorly understood.

These results represent the first structural insights on the family of CCCs and showed its dimeric organization on the purified C-terminal domain and the purified full-length transporter of MaCCC.

4 Appendix

Despite the overall promising biochemical behavior of the full-length CCC from *Methanosarcina acetivorans C2A*, no crystallization was observed even with broad screening experiments. To overcome the crystallization bottleneck two strategies were pursued, which have already given preliminary results that are described in the appendix to this work and which include attempts to crystallize a transporter lacking the cytoplasmic domain and the generation of monoclonal antibodies for the subsequent co-crystallization of the transporter in complex with Fab fragments.

4.1 Membrane Domain of Prokaryotic CCCs

4.1.1 MaCCC Membrane Domain

In an attempt to improve the possibilities for successful crystallization of the MaCCC transporter the potential flexible C-terminal domain was removed and the overexpression of the isolated membrane part of MaCCC was investigated. Different constructs of the transmembrane domain were designed taking the information from hydropathy plots and the observed boundaries from the structure determination of the C-terminal into acount (Fig 4-1).

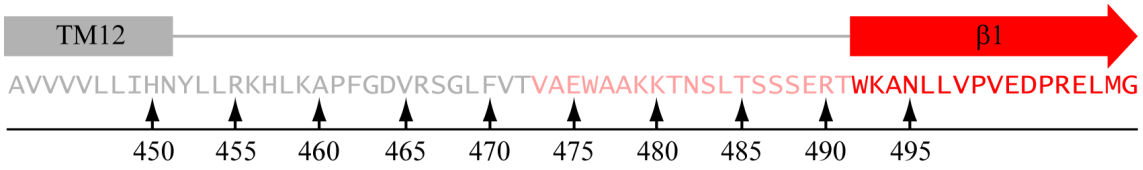


Fig 4-1 MaCCC Amino Acid Sequence of the Linker Between the Membrane Domain and the C-Terminal Domain
MaCCC membrane domain construct design: The truncation of MaCCC to remove the potentially flexible C-terminal domain was based on hydropathy analysis from the experience obtained from the structure determination of the separate cytosolic domain. A section of the MaCC sequence is shown with the last predicted transmembrane domain [TM12] and the start of the C-terminal domain structure [β 1] indicated. The construct for the expression of the C-terminal domain started at V472. The expression tests were conducted with a construct ending at residue H450 and the second trial with A473. A broad expression screen was performed with constructs varied by five amino acids in length from TM12 to β1 (arrows).

With a protein fragment ending after the last predicted transmembrane segment at residue H450 no expression was observed under various conditions with the T7 promoter using anti His-tag Western blot. A second construct considering experience from the C-terminal domain construct was created that ended at residue A473. With this construct little expression was detected by an anti His-tag immunoblot (Fig 4-2). The

low signal that points towards low expression of this construct has led to the design of various constructs with different length of the linker after the membrane domain and the C-terminal domain.

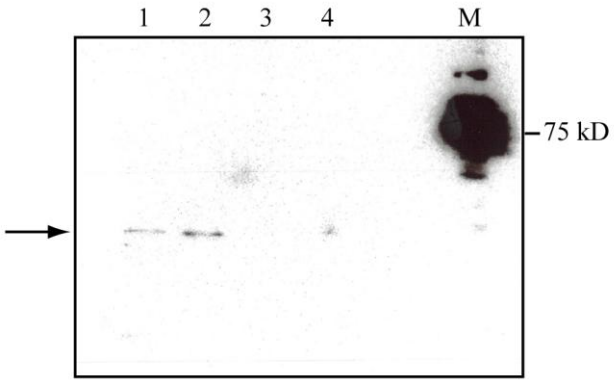


Fig 4-2 MaCCC Membrane Domain Expression Test
MaCCC was truncated after the last predicted transmembrane segment and before the start of the cytosolic domain at residue A473. After growing in TB media to an OD_{600nm}=0.5 at 37 °C the expression was induced with 0.5 mM IPTG for 3 h. Four clones (1 - 4) were tested. Two of them (clone 1 and 2) showed weak expression (arrow) that is not sufficient for purification and characterization. The signals for clone 3 and 4 were not detectable.

These expression tests were carried out with constructs, whose sequences were extended in steps of five residues after the last predicted transmembrane helix of MaCCC at residue H450 (*i.e.* residue H450) (Fig 4-1). The constructs were created using FX cloning into pBXGC3H including C-terminal GFP fusion (Fig 2-4). Three clones for each construct were investigated for their expression behavior at different arabinose concentration in a 96 well plate format by induction overnight at 25 °C and analyzed by monitoring the GFP fluorescence (Fig 4-3B). All constructs show expression with different yields. The highest expression was recorded for a construct that terminated domain at residue R490 followed by N495 and T485. These constructs have comparably long tail after the proposed end of the membrane domain.

A

	1	2	3	4	5	6	7	8	9	10	11	12
A	1-H450 #A			1-A460 #C			1-E475 #B			1-R490 #A		
B	1-H450 #B			1-V465 #A			1-E475 #C			1-R490 #B		
C	1-H450 #C			1-V465 #B			1-K480 #A			1-R490 #C		
D	1-R455 #A			1-V465 #C			1-K480 #B			1-N495 #A		
E	1-R455 #B			1-F470 #A			1-K480 #C			1-N495 #B		
F	1-R455 #C			1-F470 #B			1-T485 #A			1-N495 #C		
G	1-A460 #A			1-F470 #C			1-T485 #B			1-N495 #B	no inducti	empty
H	1-A460 #B			1-E475 #A			1-T485 #C			1-K480 #C	no inducti	empty
arabi-nose [%]	0.01	0.001	0.0001	0.01	0.001	0.0001	0.01	0.001	0.0001	0.01	0.001	0.0001

B

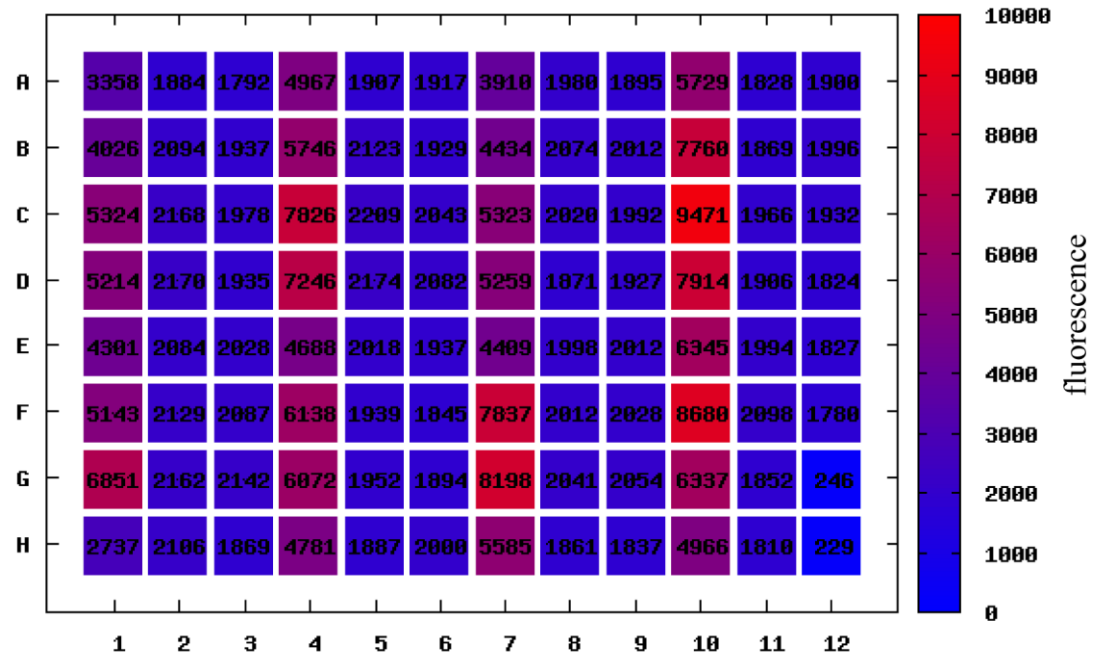


Fig 4-3 Expression Test of MaCCC Membrane Domain
The membrane domain of MaCCC was tested for overexpression with different constructs differing in the length of their C-terminal tail. **(A)** On a 96 well plate (A to H and 1 to 12) three clones (#A-C) were tested at different arabinose concentrations (0.01, 0.001, 0.0001 % (w/v)) for expression. The best expressing clones are highlighted in red. **(B)** The fluorescence of the C-terminal attached GFP was recorded with red highest expression and blue lowest expression.

The highest expressing clone for each construct was analyzed by extraction and size exclusion chromatography monitoring the fluorescence of the GFP fusion of the cell

extracts. After breakage of the cells the lysed cells were solubilized with DDM followed by high spin centrifugation. The supernatant was injected onto an analytical gel filtration column attached to an HPLC system (Agilent) and detected by measurement of GFP fluorescence (Fig 4-4). For all tested constructs a strong peak was present at a volume of 7.9 ml, except for T485#B. The retention volume of 7.9 ml corresponds to an apparent molecular weight of 120 kD. For two constructs (R490#C and N495#C) an additional peak at 7.4 ml was recorded which corresponds to an apparent weight of

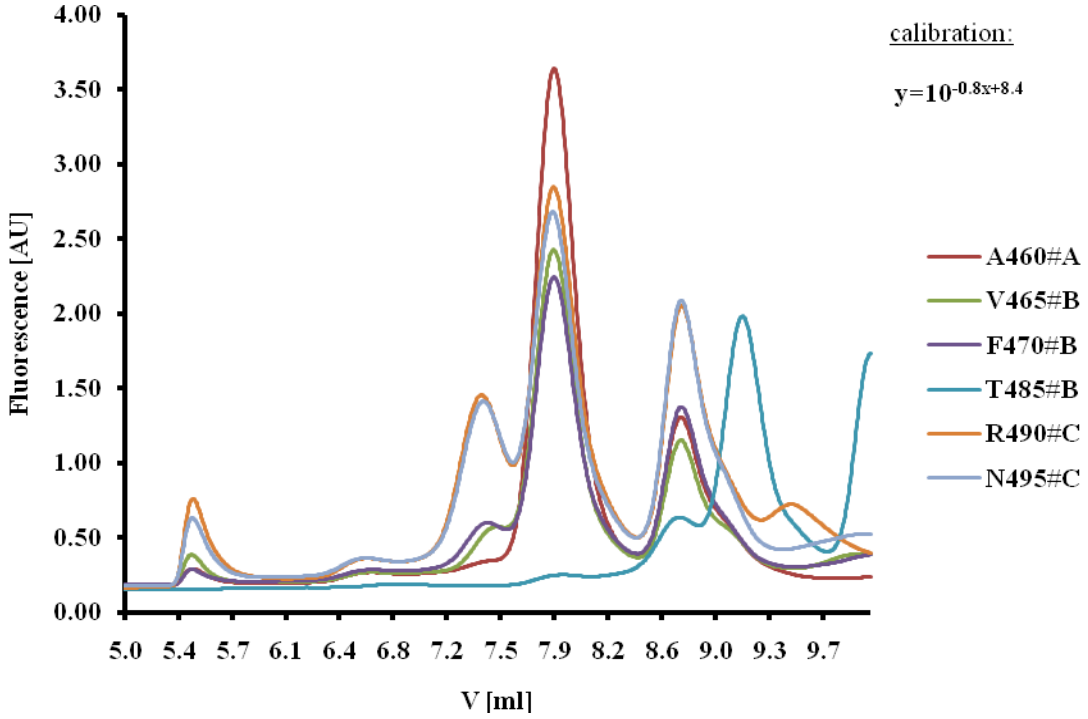


Fig 4-4 MaCCC Membrane Domain F-HPLC
The MaCCC membrane domain was analyzed by HPLC recording the fluorescence of the C-terminal attached GFP. The fluorescence in artificial units [AU] is plotted against the retention volume [ml].

300 kD. It appears that the truncated protein exists in two different oligomeric states which are influenced by increasing length of the C-terminal tail.

Future experiments for the membrane domain will include the scale-up of expression and purification of the protein. If the protein yields after scale-up are sufficient, the protein will be subjected to crystallization screening.

4.2 Antibody Production

4.2.1 Binding Molecules that Facilitate Crystallization

A strategy to support crystallization of membrane proteins is the use of soluble high affinity binders that form tight complexes with the membrane protein (Koide 2009). The

increased hydrophilic surface of the complex provides specific interactions for crystal contact formation. Since crystal contacts are predominantly mediated by hydrophilic surfaces the binding protein significantly increase the hydrophilic area of membrane proteins, whose hydrophobic membrane inserted regions are covered by detergents upon solubilization. Specific binding proteins potentially also stabilize certain conformations of a membrane protein and such allow to obtain a homogenous sample. This property can be of big importance for transport proteins that undergo large conformational changes during the transport cycle. Most common used binders in crystallography are antibody derived binders like Fabs (fragments of the antibody binding regions), Fv's (fragment of the variable region) and scFv (single chain Fv fragment). DARPins (designed ankyrin repeat proteins) present another group of soluble high affinity binders that have been explored recently. While antibodies are still predominantly derived from natural sources, after immunization of animals and by the generation of hybridoma cell s, recombinant antibody fragments selected from engineered libraries became an important alternative during the last years. DARPins in contrast are selected from large libraries by ribosome display techniques.

Fabs are built up from the variable and constant parts of the light chains and the N-terminal halve of the heavy chain of IgGs. They can be generated by partial proteolytic digest of IgGs with papain. Fv's are composed of the variable parts of heavy and light chains only and are usually linked via a short peptide linker in scFv (Fig 4-5). The

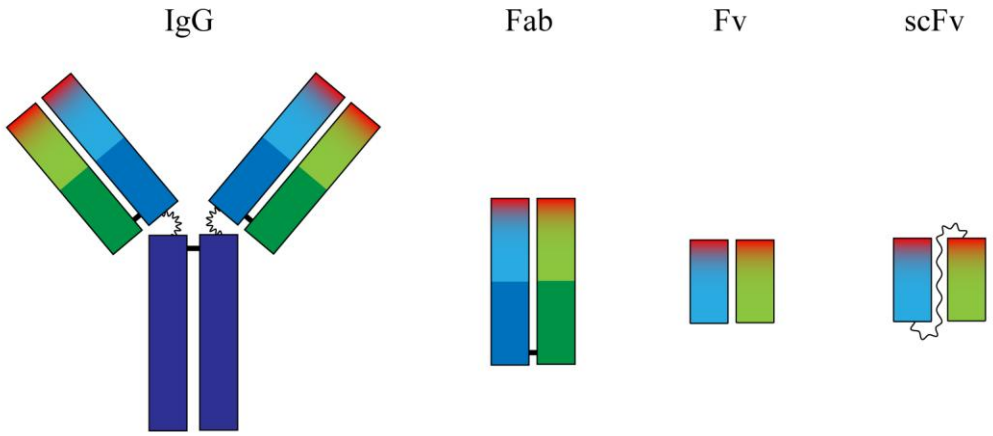


Fig 4-5 Antibody Molecules and Fragments
IgG are part of the immune system and created through hybridoma technology. They consist of two chains, heavy (blue) and light (green) chain. Interaction between the chains and several cysteine bridges (black) stabilize the structure. The complementary defining regions (CDRs) are important for antigen and are located on the tips of the arms (red). Through proteolytic cleavage in the linker region (wavy line) in the heavy chain with papain Fabs are created. For crystallization the more stiff and defined Fab-, Fv and scFv molecules are used.

chloride channel from E. coli, EcClC, was successfully crystallized with Fabs. These binders facilitated the crstallyzation and structure determination of this channel by increasing the hydrophilic surface (Fig 4-7A).

DARPins are assembled in a modular fashion that allows a high stability combined with high diversity in target binding. The general framework consensus sequence that was derived from natural ancestors is enclosed by protecting N- and C-terminal capping region to create a stable protein (Fig 4-6). A varying amount of repeats between the caps with defined randomized residues, covered by a large library allows the selection of recombinant binders for various targets.

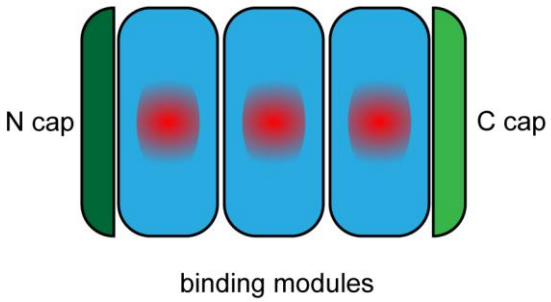


Fig 4-6 Shematic Representation of DARPins
Designed akyin repeat proteins (DARPins) are built up modular. The binding modules have a defined structure encoded by a consensus sequence. Residue varitions at restricted positions (red area) assemble the library for ribosome display. Variation in binding affinity and specificity is reached by variation of the repeat number (N=3 is shown). The N- and C-terminal caps shield the hydrophobic core and generate a stable protein.

DARPins have presently only once been successfully used in membrane protein crystallization where they were stabilizing one conformation in the transport cycle of

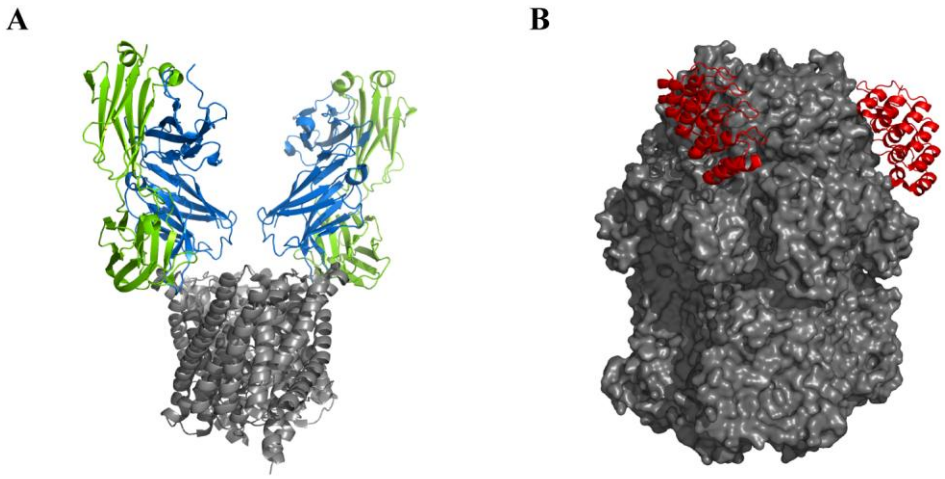


Fig 4-7 Co-Crystallization with Proteinaceous Binders
Co-crystallization with binders has various effects on the crystallization behavior. **(A)** It can dramatically increase the hydrophilic surface and thus be essential for crystal contacts like in EcClC (grey surface) with two Fabs (green and blue, light and heavy chain, respectively). **(B)** The binders can also fix a certain conformation and allow the structure determination of distinct transport states like AcrB (grey surface) in complex with two DARPins (red). EcClC (1ots); AcrB (2j8s).

the major drug efflux pump AcrB. Through selection and binding a homogenous protein species was generated that showed improved crystallization properties (Fig 4-7B). The formation of novel crystal contacts by an increase of the polar surface was in this case not observed.

4.2.2 Production and Identification of Positive Hybridoma Cell Lines

The unsuccessful crystallization screening of MaCCC motivated the production of specific binding proteins for the subsequent crystallization of complexes with the transporter. The production of monoclonal antibodies was the first choice since since this approach has already been successful in various membrane protein structure determinations. Mouse was chosen as organism for immunization with the antigen. All experiments were carried out in collaboration with CEPower, a company specialized for the generation of monoclonal antibodies in Wädenswil close to Zurich. Three mice were immunized with MaCCC in its putative monomeric conformation. The immunization was repeated twice before harvesting of the mouse spleen. The fusion of the B-lymphocytes from spleen cells of two mice with myeloma cells to form immortalized hybridoma cells expressing IgGs was carried out by standard methods and confirmed with selection in hypoxanthine aminopterin thymidine (HAT) medium. HAT medium is a selection medium where the drug aminopterin inhibits the dihydrofolate reductase, which results in an impairment of nucleotide de novo synthesis. The synthesis of thymine can be bypassed with thymidine and of guanine with hypoxanthine. Only fussed cells carrying the immortalization of the myeloma cells and the hypoxanthine-guanine phosphoribosyltransferase (HGPRT) of B-lymphocytes can survive during this selection. 497 stable hybridoma cell lines were grown and all supernatants containing secreted antibodies were screened by ELISA for a reaction against MaCCC. All ELISA signals above background were selected to have a maximum set of potential antibodies for a thorough investigation. With this procedure 51 antibody expressing clones, SW01 to SW51, were identified. These 51 clones were analyzed by ELISA on the same plate for their binding against MaCCC full-length and the C-terminal domain of MaCCC (Fig 4-8). Only few monoclonal antibodies show a strong signal against the full-length protein (SW27, 34, 35) (Fig 4-8B), or the isolated domain (SW3, 27, 28, 34, 35) (Fig 4-8C). Several antibodies appear to bind solely to the full-length protein, but with an ELISA signal only slightly above background (SW 7, 9, 12, 13, 16, 19, 20, 22, 26, 30, 33, and 49). The antibody concentration in the supernatants was estimated by coating the ELISA plate with the antibody containing hybridoma supernatant and by using an antibody against mouse IgG for the reaction. All supernatants of the selected clones SW1-51 had similar amounts of antibody but they all were very low compared to antibody amounts generated from a hybridoma cell line expressing an antibody against the chloride transporter EcClC that was used as positive control. Since the signals recorded in different ELISAs for the same sample showed a high variance, the clone SW10, 42 and 47 were added to the list of potential candidates for in-depth

characterization. Since for the initial characterization hybridoma cell supernatants containing the monoclonal antibodies were used, a next step includes the purification of antibodies and the analysis of binding against MaCCC by ELISA and by complex formation in solution after the isolation of the Fab fragments.

A

	1	2	3	4	5	6	7
A	SW1	SW9	SW17	SW25	SW33	SW41	SW49
B	SW2	SW10	SW18	SW26	SW34	SW42	SW50
C	SW3	SW11	SW19	SW27	SW35	SW43	SW51
D	SW4	SW12	SW20	SW28	SW36	SW44	empty
E	SW5	SW13	SW21	SW29	SW37	SW45	empty
F	SW6	SW14	SW22	SW30	SW38	SW46	empty
G	SW7	SW15	SW23	SW31	SW39	SW47	empty
H	SW8	SW16	SW24	SW32	SW40	SW48	empty

B C

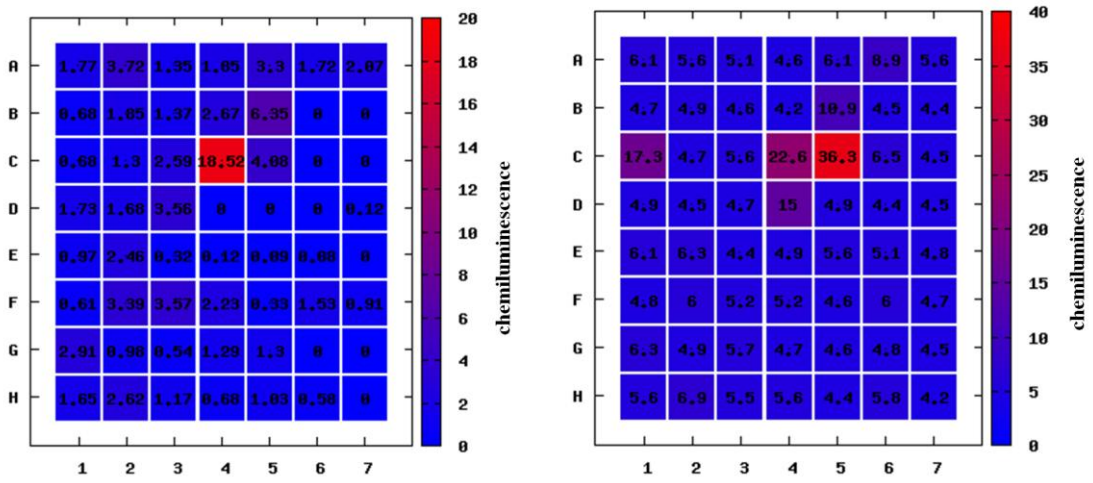


Fig 4-8 Antibody Binding Against MaCCC Full-Length and C-Terminal Domain
The selected antibodies SW1 to SW51 were tested for their binding properties to the full-length protein and to the C-terminal domain of MaCCC. **(A)** The supernatants of growing hybridoma cultures were used as primary antibody in an ELISA against **(B)** the full-length MaCCC and against **(C)** the C-terminal domain. The chemiluminescence was recorded after incubation of a secondary anti mouse antibody coupled to horse radish peroxidise.

Selected clones that showed a high siganl signal against MaCCC were characterized by Western blotting. The characterization of binding against the denatured protein by Western blot analysis allows the discrimination between sequence epitopes (which show signals in a Western blot analysis) from the structural epitopes (which do not show a signal since the structural epitope is lost on a denaturing gel). Antibodies that recognize structural epitopes are believed to be better suited for crystallization because their interaction with the antigen usually causes the formation of a rigid complex.

Sequence epitopes can be unstructured N- or C-terminal tails as well as unstructured loop regions. Purified MaCCC was separated by SDS PAGE and blotted on a PVDF

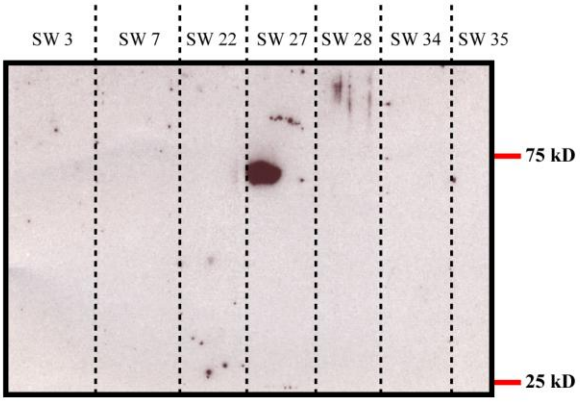


Fig 4-9 Epitope Characterization of Selected Monoclonal Antibodies
MaCCC was separated by SDS-PAGE and Western blotted. After separate incubation with the hybridoma supernatant of the indicated clone the blot lanes were developed using anti mouse antibodies and chemiluminaescene.

membrane that was subsequently blocked with milk to saturate non-specific protein binding sites. Sections of the PVDF membrane all containing MaCCC and a protein standard marker lane were incubated with the antibody containing supernatants of the selected hybridoma cells. After incubation with the secondary antibody and the detection of chemiluminescence, only one clone. SW 27, showed a signal, indicating that this antibody recognizes a sequence specific epitope. The SW27 clone also had the strongest signal in the ELISA reaction against the full-length proteins and the C-terminal domain. O clones did not show a signal in Western blot indicating that they do not recognize the disordered protein regions.

4.2.3 Fab Production and Purification

The yield of monoclonal antibody from hybridoma cell culture in cell line flasks (BD Biosciences) was low, and the amount of purified Fabs after cleaving and purification was too low for efficient and reliable testing of the antibody properties in solution. An overproduction in collaboration with CE-Power in roller bottles was more successful but still the yield was severall orders of magnitude below the yield obtained from hybridoma cell line expressing the EcClC antibody (2 - 4 mg/l for CCC antibody compared to 100 mg/l for the EcClC antibody). While the generation and purification of Fab fragments from IgG of SW 10, 22, 42, 47 and SW 49 was successful, the overall yields were very low due to various dialysis, cutting and affinity purification steps.

After the experiences with low yield of antibody overproduction with hybridoma cells another strategy was chosen to generate Fab fragments by cloning and heterologous overexpression *E. coli*. These experiments were carried out with the help of Laura Prochazka, a master student who established the cloning and expression of Fab

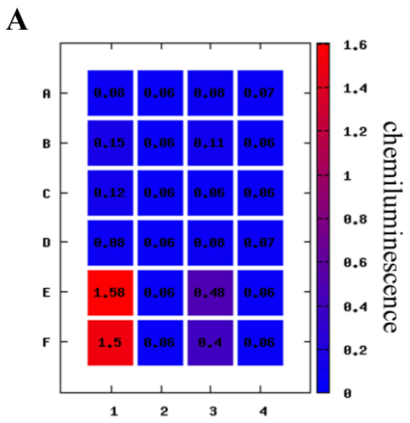
fragments. The two chains of Fab fragments were clones independently and directed to the periplasm of *E. coli*. To facilitate a high throughput analysis of the selected binders a combined FX and VBex cloning was chosen. The variable regions of both, heavy and light chains were amplified from cDNA of hybridoma cell lines. In the first FX cloning step, the variable regions of both chains were cloned in separate vectors. During cloning the variable parts were cloned on the 5' side of the vector containing the constant part of the human heavy and light chains, respectively. The second VBex step merged the two chains into a polycistronic vector containing both, heavy and light chain. The operon is controlled by the P_{LAC} promoter. Additionally both proteins start with different export signals (*phoA* for the heavy chain and *ompA* for the light chain) for periplasmic targeting and assembly of the Fabs. For later purification the heavy chain was fused to a His_6-tag on its C-terminus.

For amplification of the antibody fragments, whole cell mRNA of cultivated hybridoma cells was isolated in sufficient amounts to be translated into cDNA by reverse transcriptase with degenerated primers. The difficult amplification of the variable part of the heavy and light chain from the cDNA pool was accomplished with degenerated primer pairs designed based on of multiple antibody sequences. By this strategy Fab fragments of all selected antibodies were cloned.

For the overexpression of the Fab constructs a standard procedure was chosen. *E. coli* transformed with the expression plasmid was grown until the mid exponential state at 37 °C and induced with 0.1 mM IPTG at 25 °C overnight. After harvesting and breakage of the cells, Fabs were purified by IMAC and analyzed by SEC. Several purified Fabs show a monodisperse peak on SEC which indicates well-folded protein (Fig 4-10A).

4.2.4 Fab Binding Analysis

The purified Fabs were investigated for binding to MaCCC by ELISA in comparison to the signal of the hybridoma supernatant of the selected clones (Fig 4-11). After coating with MaCCC and incubation with the supernatant and the purified Fab the reactions were developed with anti mouse and anti His-tag antibodies, respectively. The highest signal was observed for the sequence specific clone SW27. It showed a strong signal for the hybridoma supernatant and a comparably strong signal (six times above background) for the expressed Fab. Although the supernatants of SW10 and SW12 showed also a signal above background for the hybridoma supernatant, the signal for the expressed Fab was very weak. Only for SW12 a signal above background was detected. SW35 in constrast did not show any signal for the hybridoma supernatant and the Fab.



B

	1	2	3	4
A	SW09	empty	SW09	empty
B	SW10	empty	SW10	empty
C	SW12	empty	SW12	empty
D	SW35	empty	SW35	empty
E	SW27	empty	SW27	empty
F	SW27	empty	SW27	empty
	hybridoma supernatant	anti mouse	purified Fab	anti His-tag

Fig 4-11 Anti-MaCC Fab Binding Test
Comparative ELISA analysis of hybridoma supernatant and purified Fab binding to MaCCC. **(A)** The chemiluminescence was recorded after coating with MaCCC and incubation with the samples and development with HRP. **(B)** Column 1 represents the hybridoma supernatant samples that were detected with anti mouse antibodies while in column 3 the purified Fabs were tested with detection through the His-tag on the heavy chain with an anti His-tag antibody. Column 2 and 4 are the background signals for the detection methods.

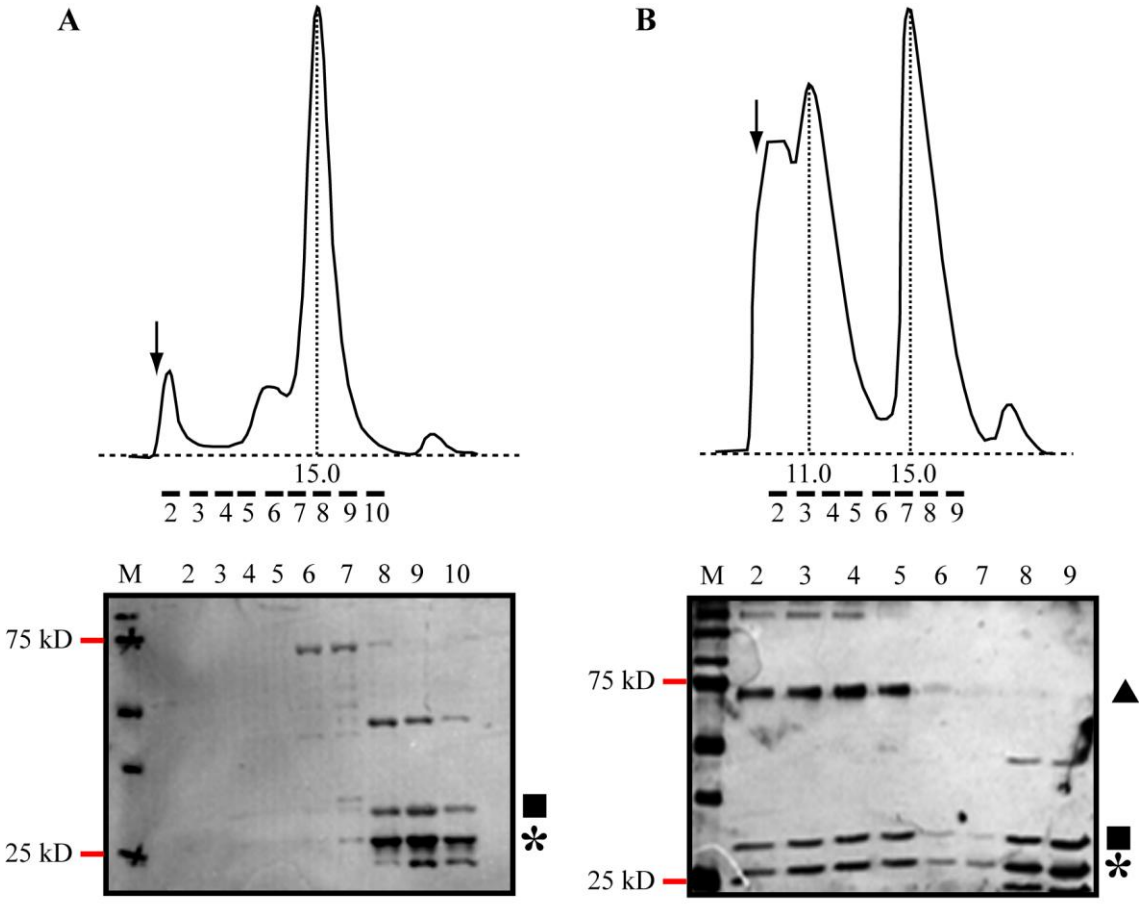


Fig 4-10 Purification and Binding of SW27 Fab
(A) SW27 Fab with a C-terminal His-tag on the heavy chain was purified using IMAC and SEC. The SEC fractions were analyzed by Western blot with anti His-tag antibody (black quadrate) and anti mouse light chain antibody (asterisk) and overlaid. **(B)** Binding of SW27 Fab to MaCCC was proven by co-migration on SEC and silver stain of the fractions after SDS-PAGE. (quadrate-heavy chain, asterisk-light chain, triangle-MaCCC).

To investigate the formation of a stable complex MaCCC was incubated with the

respective Fab fragment for 30 min at 4 °C at a ratio of 3 mol Fab to 1 mol MaCCC. Since specific Fab fragments usually have a very high affinity in the nano-molar range at three-times excess of Fabs was expected to be sufficient for a quantitative complex formation. The complex was investigated by SEC to observe the coelution of MaCCC and Fabs and to separate unbound Fab from the complex. Since SW27 showed strongest binding to MaCCC it was selected for the binding experiment (Fig 4-10B). In the SEC experiment the peak of MaCCC with an elution volume of 11.9 ml was shifted to an elution of 11 ml, due to the binding of the SW27 Fab as shown by SDS-PAGE of the SEC fractions.

Unfortunately SW27 proves to be the only Fab that showed binding to MaCCC in solution (Tab 4-1) whereas other Fabs that also had low signals on ELISA did not bind to MaCCC.

Clone	**Source**	**Fab binding**
SW9	*E. coli*	No
SW10	hybridoma cells/ *E. coli*	No/(ELISA)
SW12	*E. coli*	No
SW22	hybridoma cells	No
SW27	*E. coli*	ELISA and SEC
SW28	*E. coli*	No
SW35	*E. coli*	No
SW42	hybridoma cells	No
SW47	hybridoma cells	No
SW49	hybridoma cells	No

Tab 4-1 Fab Binding to MaCCC

4.2.5 Materials and Methods

96 Well Expression Test. The truncated constructs for the MaCCC membrane domain were cloned into pBXGC3H vector (Fig 2-5B) and three clones were picked from each construct for overnight preculture in LB medium at 37 °C. A 96 2.2 ml deep well plate (round bottom, Thermo Scientific) was filled with 1 ml TB medium and inoculated with 1 % preculture and closed with gas permeable seals (Thermo

Scientific). After growth at 37 °C with vigorous shaking until OD_{600nm}=0.5 (~135 min) the cells were immediately cooled to 25 °C and induced with 10^{-3} (w/v) arabinose. After overnight incubation the cells were centrifuged and the OD_{600nm} was determinate by resuspension of the cell pellet in 100 µl PBS and measurement of an at least 1:9 dilution to 100 µl in a MaxiSorp (Nunc) 96 well plate. Before the resuspended cells were analyzed for fluorescence ($\lambda_{absorption}$=485 nm, $\lambda_{emission}$=535 nm) in a 96 black well plate (Nunc) with a Tecan GENios system and with the Magellan software. After addition of 300 µl ice-cold lysis buffer (50 mM KP_i (pH 7.2), 150 mM NaCl, 1 mM $MgSO_4$, 10 % (w/v) glycerol, 1 mM PMSF and 20 µg/ml DNase) and 300 mg glass beads (0.1 mm diameter) the samples were broken in a FastPrep device (Bio101) for two times 20 s at force 6 with 5 min cooling on ice between the runs. 100 µl of the lysed cell suspension was incubated with 25 µl 10 % DDM for 1 h at 4 °C and high-spin centrifuged (30 min, 4 °C, x g). 10 to 30 µl were injected onto a Zorbax GF-250 column (Agilent) using a HPLC system with fluorescence detector (8 µl FLD cell, Agilent Technologies 1200 Series).

ELISA. 96 well MaxiSorp plates (Nunc) were coated with 100 µl of 100 nM protein in buffer B (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM DM) and incubated for 2 h at 4 °C. After washing 3 times for 5 min with 200 µl buffer B the plate was blocked for 1 h at with 300 µl buffer B + 5 % BSA (w/v) and subsequently washed. After incubation with 100 µl hybridoma supernatants (1:1 mixed with 2 x buffer B with 4 % BSA (w/v)) for 1 h at and washing, the plate was incubated for 1 h at room temperature with 100 µl goat-antimouse-HRP antibody diluted in buffer B + 2 % BSA (w/v) for 1 h. After washing the plate was developed with 100 µl prepared TMB substrate (Sigma) and stopped with 100 µl 1 M HCl. The signal was measured with absorbance at 450nm. All steps were carried out at room temperature if not stated different.

Immunization of Mice and Fusion to Myeloma cells. The immunization was carried out by CE Power only the target protein was prepared and supplied. Three mice were immunized with 300 µl of 200 µg/ml target protein. This procedure was repeated two times and 10 days after the last immunization the spleen cells were harvested. The spleen B-cells were fused to myeloma cells (ΔHGPRT) and cloned with limited dilution. The supernatants were tested for positive clones by ELISA.

Expression of IgGs. Hybridoma cells were thawed and after removal of DMSO by centrifugation and resuspension in hybridoma medium supplemented with 5 % fetal calf serum (FCS) and 2 % L-glutamine the cells were grown with at least two passages until cell densities of 2 x 10^6/ml. The CELLine 1000 device (BD Biosciences) was prewetted and 15 ml of high density hybridoma cells were filled into the cell compartment followed by filling 1 l of prewarmed medium in the nutrient compartment. After incubation for seven days the cell were aspirated using a serological pipet, cenrifugated

with keeping the IgG containing supernatant and the cell pellet was diluted to reach the mentioned density and refilled into the cell compartment. This procedure can be repeated twice and after 21 days of growth the feeding medium has to be replaced.

Preparation of Fabs from IgGs. The hybridoma supernatant was dialyzed against overnight (MWCO=8000 Da). MonoQ sepharose fast flow (Pharmacia) was equilibrated with dialysis buffer and the sample was loaded. The protein was eluted with a gradient from 10 mM NaCl to 400 mM NaCl in 10 mm Tris-HCl pH 8.0. The first peak was collected and dialyzed overnight against PBS and the antibody diluted to OD_{280nm}=1.5 The cutting was performed after addition of 10 mM β-mercaptothanol, 10 mM L-cysteine, 10 mM EDTA (pH 7.0) and incubation for 30 min with the addition of papain at a ratio 1:100, papain:IgG. After completion of the cutting the reaction was stopped with 23 mM iodoacetamine for 20 min at 37 °C. After dialysis overnight (MWCO=1000 Da) against 10 mM Tris-HCl pH 8.0, 10 mM NaCl the Fabs and Fcs were separated by MonoQ sepharose. Fabs were in the flow-through.

Cloning of Fabs. According to Master thesis Laura Prochazka, Zurich 2009

5 Bibliography

A. J. McCoy, R. W. G.-K., P. D. Adams, M. D. Winn, L.C. Storoni and R.J. Read. (2007). "Phaser crystallographic software." J. Appl. Cryst. **40**: 658-674

Abramson, J. and E. M. Wright (2009). "Structure and function of Na+-symporters with inverted repeats." Curr Opin Struct Biol **19**(4): 425-432.

Accardi, A. and C. Miller (2004). "Secondary active transport mediated by a prokaryotic homologue of ClC Cl- channels." Nature **427**(6977): 803-807.

Adams, P. D., R. W. Grosse-Kunstleve, et al. (2002). "PHENIX: building new software for automated crystallographic structure determination." Acta Crystallogr D Biol Crystallogr **58**(Pt 11): 1948-54.

Anselmo, A. N., S. Earnest, et al. (2006). "WNK1 and OSR1 regulate the Na+, K+, 2Clâˆ’ cotransporter in HeLa cells." Proceedings of the National Academy of Sciences **103**(29): 10883-10888.

Bernard, P., P. Gabant, et al. (1994). "Positive-selection vectors using the F plasmid ccdB killer gene." Gene **148**(1): 71-4.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 905-21.

CCP4 (1994). "Collaborative Computational Project Nr. 4. The CCP4 Suite: Programs for X-ray crystallography." Acta Crystallogr. D **50**: 760-763.

Cowtan, K. (1994). "An automated procedure for phase improvement by density modification." Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography **31**: 34-38.

de La Fortelle, E. and G. Bricogne (1997). Methods in Enzymology. Methods in Enzymology. C. W. Carter and R. M. Sweet. New York, Academic**:** 492-494.

Dutzler, R., E. B. Campbell, et al. (2002). "X-ray structure of a ClC chloride channel at 3.0 A reveals the molecular basis of anion selectivity." Nature **415**(6869): 287-94.

Gamba, G. (2005). "Molecular physiology and pathophysiology of electroneutral cation-chloride cotransporters." Physiol Rev **85**(2): 423-93.

Geertsma, E. R., M. Groeneveld, et al. (2008). "Quality control of overexpressed membrane proteins." Proceedings of the National Academy of Sciences **105**(15): 5722-5727.

Geertsma, E. R. and B. Poolman (2007). "High-throughput cloning and expression in recalcitrant bacteria." Nat Meth **4**(9): 705-707.

Gimenez, I. and B. Forbush (2005). "Regulatory phosphorylation sites in the NH2 terminus of the renal Na-K-Cl cotransporter (NKCC2)." Am J Physiol Renal Physiol **289**(6): F1341-5.

Granseth, E., S. SeppÃ¤lÃ¤, et al. (2007). "Membrane protein structural biology â€“ How far can the bugs take us? (Review)." Molecular Membrane Biology **24**(5): 329 - 332.

Isenring, P. and B. Forbush (2001). "Ion transport and ligand binding by the Na-K-Cl cotransporter, structure-function studies." Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology **130**(3): 487-497.

Johansson, L. C., A. B. Wöhri, et al. (2009). "Membrane protein crystallization from lipidic phases." Current Opinion in Structural Biology **19**(4): 372-378.

Kabsch, W. (1993). "Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown

Symmetry and Cell Constants." J. Appl. Cryst. **26**(6): 795-800.

Kawate, T. and E. Gouaux (2006). "Fluorescence-Detection Size-Exclusion Chromatography for Precrystallization Screening of Integral Membrane Proteins." **14**(4): 673-681.

Koide, S. (2009). "Engineering of recombinant crystallization chaperones." Curr Opin Struct Biol **19**(4): 449-57.

MacCallum, R. M. (2004). "Striped sheets and protein contact prediction." Bioinformatics **20 Suppl 1**: i224-31.

Mercado, A., V. Broumand, et al. (2006). "A C-terminal domain in KCC2 confers constitutive K+-Cl- cotransport." J Biol Chem **281**(2): 1016-26.

Moore-Hoon, M. L. and R. J. Turner (2000). "The structural unit of the secretory Na+-K+-2Cl- cotransporter (NKCC1) is a homodimer." Biochemistry **39**(13): 3718-24.

Morth, J. P., B. P. Pedersen, et al. (2007). "Crystal structure of the sodium-potassium pump." Nature **450**(7172): 1043-1049.

Oldham, M. L., A. L. Davidson, et al. (2008). "Structural insights into ABC transporter mechanism." Current Opinion in Structural Biology **18**(6): 726-733.

Olesen, C., M. Picard, et al. (2007). "The structural basis of calcium transport by the calcium pump." Nature **450**(7172): 1036-1042.

Otwinowski, Z. and W. Minor (1997). "Processing of X-ray diffraction data collected in oscillation mode." Methods Enzymol. **267**: 307-326.

Pape, T. and T. R. Schneider (2004). "HKL2MAP: a graphical user interface for phasing with SHELX programs." J. Appl. Cryst. **37**: 843-844.

Parvin, M. N., T. Gerelsaikhan, et al. (2007). "Regions in the Cytosolic C-Terminus of the Secretory Na+-K+-2Cl- Cotransporter NKCC1 Are Required for Its Homodimerizationâ€ " Biochemistry **46**(33): 9630-9637.

Sanner, M. F., A. J. Olson, et al. (1996). "Reduced surface: an efficient way to compute molecular surfaces." Biopolymers **38**(3): 305-20.

Schneider, T. R. and G. M. Sheldrick (2002). "Substructure solution with SHELXD." Acta Crystallogr D Biol Crystallogr **58**: 1772-1779.

Schomberg, S. L., J. Bauer, et al. (2003). "Cross Talk Between the GABAA Receptor and the Na-K-Cl Cotransporter Is Mediated by Intracellular Cl." J Neurophysiol **89**(1): 159-167.

Schuck, P. (2000). "Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling." **78**(3): 1606-1619.

Schuck, P. (2000). "Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling." Biophys J **78**(3): 1606-19.

Schuck, P., M. A. Perugini, et al. (2002). "Size-Distribution Analysis of Proteins by Analytical Ultracentrifugation: Strategies and Application to Model Systems." Biophys. J. **82**(2): 1096-1111.

Slotboom, D. J., R. H. Duurkens, et al. (2008). "Static light scattering to characterize membrane proteins in detergent solution." Methods **46**(2): 73-82.

Thompson, J. D., D. G. Higgins, et al. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-80.

Curriculum Vitae

Stefan Gerhard Josef WARMUTH

Date of Birth	11th of May 1980
Place of Birth	Ludwigslust, Germany
Nationality	German

Education

2004-2009	Phd thesis at the University of Zurich and Molecular Life Science PhD Program under the supervision of Prof. R. Dutzler
2004	Diploma thesis at the Insitute of Microbiology, ETH Zurich; Title: "c-Ring der A-Typ ATPase aus *Methanopyrus kandleri*: Heterologe Expression in *E. coli* und Reinigung aus Wildtyp Zellen" under the supervision of Prof. P. Dimroth
2002-2004	Studiengang Biochemie und Molekularbiologie an der ETH Zürich; Abschluss: Diplom
1999-2002	Studiengang Biochemie an der Eberhard-Karls-Universität Tübingen; Abschluss: Vordiplom
1990-1998	Heinrich-Schliemann-Gymnasium Schwerin; Abschluss: Abitur

Professional Development

2008	Talk - Gordon Research Conference „Membrane Transport Proteins" Lucca, Italy; Title: "The Structure of the C-terminal Domain of a Cation Chloride Cotransporter"
2008	Poster Presentation – Gordon Research Conference „Membrane Transport Proteins" Lucca, Italy; Title: "X-ray Structure of the Cytoplasmic Domain of a Prokaryotic Cation Chloride Cotransporter"

2008	Poster Presentation – NCCR Symposium, Zurich, Switzerland; Tilte: "X-ray Structure of the Cytoplasmic Domain of a Prokaryotic Cation Chloride Cotransporter"
2005	Postgraduate Course – Swiss Light Source at the Paul Scherrer Institute, Villingen, Switzerland; Title: "Synchrotron Data Acquisition Techniques in Macromolecular Crystallography"

Publication

Warmuth, S., Zimmermann, I. & Dutzler, R. X-ray Stucture of the C-terminal Domain of a Prokaryotic Cation-Chloride Cotransporter. Structure 17, 538-546 (2009)